THE VALIDATION OF SEASHORE'S MEASURES OF MUSICAL TALENT BY FACTORIAL METHODS

By JOHN McLEISH
Extra-Mural Department, University of Leeds

I. *Problem.* II. *Subjects and Tests.* III. *Results :* (a) *Factor Analysis of the Seashore Battery ;* (b) *Multiple Correlations with External Criteria.* IV. *Summary and Conclusions.*

I. PROBLEM

The object of the research described in the following paper was to attempt to validate the Seashore tests of musical ability by a factorial procedure, using both internal and external criteria. The need for both kinds of validation appears to be now generally admitted in the more familiar fields of intelligence and educational testing. But it has been almost wholly ignored in the construction of tests for musical ability.

The problem of scientific validation is one of the most difficult with which the constructor of any kind of test has to deal. The commonest practice is to take some external criterion with which the test-results are then compared. This, however, is logically unsatisfactory, since it is just because external evidence is tacitly assumed to be unreliable that we have undertaken to construct a standardized test. The argument on which the validation is based is consequently circular : a criterion, admittedly imperfect, is used to demonstrate the worth of the test which is to replace it. The use of multiple criteria does not altogether release us from the logical dilemma to which we are thus reduced, for the multiplication of dubious or biased testimony does not necessarily make the conclusion more certain.

The argument may perhaps best be illustrated from the field of intelligence testing. In his earlier work, Spearman used as his criterion the ratings for intelligence submitted by teachers. He believed that the unreliability of their judgments could be eliminated by securing multiple criteria, i.e., ratings from two or more teachers for the same children, and then correcting the raw correlation between the average judgments of the teacher and the various tests he used. But, as Burt showed in his 1909 article,[1] the judgments of most teachers have a bias towards verbal ability and ability to learn ; he therefore proposed instead to use an *internal* criterion in the shape of the ' highest common factor ' underlying the whole table of correlations between the tests selected according to a psychological plan. At the same time, in deciding the nature of the internal factor so found, as he went on to insist, it was not sufficient to trust to the names of the tests or to subjective notions of what they are intended to test ; an objective check, based on improved *external* evidence, is always desirable. The process is then, not a vicious circle, but rather a spiral of successive approximation.

The danger of omitting this double check becomes especially clear in the endeavours made from time to time to validate tests of musical ability, such as the Seashore battery (9). Dr. Stanton's ' Eastman experiment ' is perhaps the most comprehensive and painstaking attempt at validation of any musical test so far carried out. The methods used in that experiment (11) and in other investigations by the same author (12, 13) included retests of students of music after training, comparison of their scores in the tests with their teachers' ratings, the development of a cumulative ' key for prognosis ' by empirical studies, a comparison of the gradings furnished by this key with all other data available about the students—their class-marks, recitals, dismissals, etc. On the basis of the results, Dr. Stanton concludes that the ' cumulative key ' (which included both the Seashore records and an intelligence test) forms a " useful administrative tool for the selection and classification of students ; . . . the tests, which became the basis of classification, can do quickly what the faculty could learn only after long observation and experience with the students " (11).

At the same time she fully recognizes that the manifestation of musical talent depends on many circumstances apart from the possession of the innate ability (or set of abilities) which the Seashore

[1] ' Experimental Tests of General Intelligence,' *Brit. J. Psychol.*, III (1909), pp. 168f.

tests were designed to measure. Among the more obvious of the extraneous influences that play an important part in determining the level of performance or appreciation, she enumerates the following : " poor health, insufficient finances, lack of ambition or effort, family demands, personality problems, religious conflicts, ' inferiority complexes,' insufficient emotional poise, misdirected goals, disintegrated concepts of prerogatives [*sic*], incoherent plans, unfavourable rapport between teacher and pupil, poor musical training, inadequate musical preparation for the educational level attempted, extreme personal jealousies, marriage, over-anxious parents, exploitation by patrons, teachers, or parents " (11, p. 131). Such a list is, in effect, a list of flaws in the normal validating procedure.

The investigation reported in the following pages should be regarded as a supplement to Dr. Stanton's investigations rather than a criticism of them. Its chief aim was to make good the more obvious shortcomings of previous enquiries, namely, the failure to apply available statistical techniques and in particular the techniques of factor analysis.

The earliest mention of ' factors ' underlying musical ability, which is based on any formal statistical procedure, appears to be that contained in the Board of Education's *Report on Psychological Tests of Educable Capacity* (introductory chapter, prepared at the request of the Committee by Burt, p. 20) ; and in his survey of group factors, published a year or two later (2), he mentions " a distinct musical capacity, no doubt highly complex," as definitely established. The data on which this conclusion was based were obtained from experiments carried out by Burt and Pelling for purposes of educational and vocational guidance, partly in L.C.C. schools and partly at the National Institute of Industrial Psychology. The novel feature in the approach consists in the addition of a number of more complex test-items, intended to measure " the Gestalt-qualities of musical understanding, as contrasted with the atomistic approach adopted by Seashore and most other investigators hitherto." [1] Attempts were later made by a succession of research-students, working at the Psychological Laboratory, University College (17 and refs.), to determine the more important components of this ' highly complex ' capacity : the factors found included elementary perceptual abilities both auditory and kinæsthetic, auditory and kinæsthetic imagery, a broad factor for æsthetic appreciation generally, a more specialized factor for musical appreciation as such (largely influenced, it would seem, both by musical experience and by temperamental inclinations), and finally general intelligence.[2] The last and the most recent of these researches was that carried out by Wing (18 : for fuller details reference should be made to his long and elaborate thesis (20), which also gives a fairly complete historical review).

Several writers have published the inter-correlations between the tests in the Seashore battery ; Brown (1) and Drake (4) have published figures which seem fairly representative ; and Wing (18, 19) reports the results of a factor analysis of the Seashore tests by Burt. With the exception of the last, which is admittedly very tentative in its nature and limited in its scope, none of those who have attempted to assess the validity and reliability of the battery have taken the logical step of factorizing the tables of correlations between its constituent tests. Instead, both Seashore's disciples and his critics seem often content to rest their conclusions largely on anecdotes of personal impressions. As a result, it is impossible to reconcile the views of American investigators like Stanton with the contrary statements of critics like Wing, who contends that the tests are atomistic, boring, musically meaningless, irrelevant, and invalid (21).

The methods of factor analysis would seem to be pre-eminently adapted for the solution of the logical difficulties which arise in studying such batteries. It is unfortunate, however, that the objectivity which results from the use of such techniques is so often lost by the way in which the factors are subsequently identified. This, as we have seen, often rests solely on subjective interpretation. Even those who have claimed to demonstrate a factor of musical ability underlying their tests have either used no external check at all or else have been satisfied to compare the test-results with ratings obtained from teachers, whose opportunities for judging genuine musical ability are, as a rule, extremely limited : ability in playing the piano or singing in tune is hardly a safe criterion.

[1] Burt, C., ' The Psychology of Music ' (Gresham Lectures, 1932, unpublished, cited by Wing (18), p. 83) ; also *id.*, ' The Factorial Analysis of Emotional Traits,' *Char. and Pers.*, VII, 1939, pp. 293–9.

[2] 17, pp. 266f., and refs. The tests used by Burt and his collaborators consisted, not only of tests of the more elementary processes—discrimination of pitch, rhythm, loudness, etc.—but also of appreciation of melody, harmony, short musical passages, longer extracts, and ' altered versions ' (given partly by the piano and partly by gramophone records). Paper-and-pencil tests were also included, e.g., marking well-known tunes in order of preference for a concert programme.

In the light of the criticisms expressed by Burt and his students—Wing, Williams, and others—it was felt that a new approach to the question of validation of the battery was needed. In particular, Seashore's assertions that intelligence and other non-musical factors had little or no influence on these measures of ' elemental ' musical ability required to be examined in the light of experimental results. Part of the aim of the present investigation therefore was to attempt to establish whether the factors resulting from an analysis of the correlations between the several tests could be approached more objectively by comparing the factors with independent evidence regarding other musical and non-musical abilities of the persons tested. As a result it is claimed that a closer approximation to an objective interpretation has been achieved.

The question that presents itself at the outset may be formulated as follows. What elements in the testing situation, apart from differences in musical abilities of the subjects, might conceivably affect the scores obtained ? A preliminary study of the test battery and the relevant literature suggested that the following elements were probably important (8, 10) :

(i) *Age.* In view of Seashore's expressed opinion about his ' measures of musical talent '—that pitch discrimination (8, p. 65), sense of time (p. 113), consonance (p. 157), tonal memory (p. 242), intensity (p. 99), and even the sense of rhythm (p. 126) are innate *elemental* capacities not affected to any considerable extent by training or environment—it is evidently desirable to ascertain the actual relations existing between the age of the subjects tested and their scores on the six tests in the battery. Seashore holds that these six capacities mature quickly in very early childhood ; so that by the time the child sufficiently understands the instructions it is possible to obtain a quantitative measurement of his natural capacities. On the other hand, Stanton and Koerth (12) have found marked differences between the performances of pre-adolescent, adolescent, and post-adolescent groups, which obviously throws doubt on Seashore's assertion. They explain the discrepancy as arising from the fact that, in the case of the adult, one is actually measuring the *psycho-physical* threshold, whereas with the child one is measuring the *cognitive* threshold. The former, they hold, cannot be measured until the age of maturity, which they place at sixteen years. Unfortunately they report no correlations between age and score for the various groups studied.

(ii) *Intelligence.* Seashore, as we have seen, claims that intelligence has very little influence on the scores obtained with his battery, provided, of course, that the subject possesses the necessary minimum of intelligence required to obey the instructions (8, pp. 56, 98, 113, 157). On the other hand, Burt (quoted by Wing, 21, p. 342) found that " when intelligence has been partialled out, the net correlations between the tests drops to little over 0·10, and are thus, with the groups we have tested, statistically insignificant." He concludes that most of the correlation between the tests, at any rate with children of the type he tested (namely, pupils at elementary schools), is due quite as much to intelligence as to innate musical ability. With older groups, more homogeneous in general intelligence, for example, children of a secondary (grammar) school type or students, the effect of intelligence proved to be somewhat less important.

(iii) *Speed of Judgment.* On introspective grounds, speed of judgment would seem to be an influence that might have quite a considerable weight in affecting the scores. With the tests of Memory, Rhythm, and Consonance, although the stimuli are presented at a speed suitable for distinct hearing, nevertheless the test itself requires a definite decision : it is a matter, not of just listening, but also of making up one's mind about small points of difference ; and at the rate the stimuli are given mere quickness of decision would seem to be important.

(iv) *Interest and Boredom.* As previously remarked, it has been suggested by Burt and Wing (21) that, for persons who are interested in music, the lack of a genuinely *musical* content renders the Seashore tests comparatively meaningless and boring. This, they say, becomes most conspicuous with the professional musician. Seashore noted much the same reaction in certain musicians he tested ; however, on retesting, after taking care to establish better rapport and eliminate the attitude of negativism, the difficulty, he claims, was overcome, and the correct psycho-physical level of the subjects secured (10).

(v) *Musical Training.* Seashore (8, 10) (and both Stanton and Koerth confirm his opinion (12, 13)) claims that the scores of subjects are uninfluenced by the amount of musical training they have had before the test. According to Seashore, the tests are truly elemental, that is, " so simple and natural that the resulting record does not vary with training " (8, p. 65).

In view of these and other considerations, the following problems were formulated for experimental study :—

(*a*) Does the Seashore battery measure a unitary ability, such as would be manifested by a general factor, or does it measure different and independent aspects of hearing, as Seashore himself suggests?

(*b*) Assuming that the battery does measure a unitary ability, is it possible to say in what this consists? Alternatively, if there is a 'general factor,' is it conceivable that it might represent not a unitary ability, but an aggregate of many common elements?

(*c*) In the latter case, assuming that these common elements can be statistically eliminated e.g., by partial correlation, and that there is still a residuum left over and above these, can this residuum be identified with 'musical talent'?

(*d*) Lastly, since Wing has been the most consistent critic of the Seashore tests, how far do the scores obtained with the Seashore battery by a given group of subjects agree with their scores for Wing's own tests?

It will be seen that this approach to the problem of validation is in direct contrast to Seashore's own attitude. His procedure can be described as basically *ad hoc*. He conceives of the 'measures' primarily as a sieve for the elimination of the musically unfit, rather than as a quantitative instrument for measuring musical ability. His use of the psychogram in place of weighted or unweighted averages for describing individual endowment is the natural corollary of his view that the abilities measured are independent. There seems in fact, as we have already implied, to be a striking similarity between Seashore's conceptions and methods and those of intelligence testers of the pre-factor era. Seashore's supporters might perhaps retort that the later development of a mathematical superstructure has tended to make its employers lose sight of significant individual peculiarities behind a mass of numerical coefficients. In Seashore's statistical methods one never forgets the individual; but, on the other hand, no clear psychological picture of the abilities tested is developed.

II. SUBJECTS AND TESTS

The tests described below were applied to a group of 100 subjects, all undergraduate or postgraduate students of psychology. To ensure comparable conditions in the procedure, every test was administered and scored by the writer. The experiments were treated as part of a normal laboratory course in experimental psychology, and in this way maximum co-operation was ensured. The series of experiments was presented as a research project, the reasons for each variable in the tests and methods being explained in advance, so as to arouse interest, but always of course so as not to interfere with the proper application of the tests. As a result, interest and motivation were very high. The two or three students who at first appeared uninterested were afterwards found to be 'tone-deaf' or to have an exceptionally low degree of musical sensitivity. In all, the following 22 variables were included in the enquiry.

1. *Age.*

2. *Intelligence* (measured by the Cattell Group Test IIIA, (*a*) timed (for 'speed'), (*b*) untimed (for 'power').

3. *Speed of reaction.* For the purpose of measuring this variable, a cancellation test was used. A reaction time test, with an auditory stimulus, was impracticable with so large a group.

4. *The Oregon test of musical discrimination.* This was scored in two parts: (*a*) correct preference for the original musical passage; (*b*) recognition of the nature of the alterations made in the original piece (i.e., whether the alteration is of melody, harmony, or rhythm).

5. *Recognition of orchestral instruments.* This test was included in order to ascertain the amount of informal musical training or interest possessed by each subject. It consisted of fifteen instruments playing in succession the same passage from Schubert's *Ave Maria*, the instruments being recorded on a Wirek recorder in random order. The subjects were asked to name the instrument, or, failing this, to say to which section of the orchestra it belonged.

6. *The Seashore measures of musical talent.* These consist of measures of pitch discrimination, intensity, time, rhythm, consonance, and tonal memory (9). Seashore's procedure was followed, except that in the case of pitch, intensity, and time, when the subject could detect no difference, he was asked to reply 'equal,' instead of guessing.

7. *Rating for interest or boredom.* Immediately after the application of each musical test, the subjects rated it for interest on a five-point scale (20). Interest was broadly described as the ease with

which attention could be maintained throughout the test, and the various points on the scale were carefully defined.

8. *Wing's test of musical intelligence.* This consists of the seven tests—chord analysis, pitch change, memory, rhythmic accent, harmony, intensity, phrasing—presented, like the Seashore battery, on gramophone records (21). Unfortunately, only half of the subjects completed all the seven tests in the battery ; the rest completed only the first five. Nearly all the subjects found some difficulty in sustaining attention, even though two sessions were allowed with rest pauses between. This perhaps is partly attributable to the recording, and partly to the use of the piano for every test.

9. *Questionnaire on musical interest and training.* A questionnaire of the open or essay type (15) was freely adapted from the questionnaire used by Seashore (8, p. 17). The subjects were encouraged to write at length on their interests in music, and to be as specific as possible in mentioning composers and pieces.

It should be emphasized that the methods adopted were a compromise between the possible and the desirable. In particular, the cancellation test, the interest ratings, and the questionnaire were included only as rough trials ; and the conclusions base on them must be regarded as suggestive rather than conclusive.

III. RESULTS

In this paper only those results which have a bearing on the validity of the Seashore battery will be reported. In the calculation of data, all scores were first normalized and standardized. Calculations were taken to four places of decimals, but only two places are printed.

(*a*) *Factor Analysis of the Seashore Battery.* The correlations obtained are shown in Table I. With one non-significant exception they are all positive ; and range from little over zero to 0·43. Except four (all correlations with Rhythm) every one is significant.

TABLE I. INTER-CORRELATIONS OF THE SEASHORE TESTS

	Pitch	Intensity	Consonance	Memory	Time	Rhythm
Pitch	(·85)	·39	·27	·42	·25	·14
Intensity ..	*·05*	(·89)	·18	·22	·28	·10
Consonance ..	*·06*	*·06*	(·34)	·43	·24	—·12
Memory ..	*·05*	*·06*	*·05*	(·93)	·23	·43
Time	*·06*	*·06*	*·06*	*·06*	(·81)	·03
Rhythm ..	*·06*	*·07*	*·07*	*·05*	*·07*	(·82)
Factor I ..	·62	·48	·39	·87	·42	·28
Factor II ..	·18	·27	·16	—·42	·26	—·45

Split-half reliabilities are shown in the diagonal cells, and probable errors below the diagonal.

The coefficients are of the same order of magnitude as those reported by Brown (1) ; while the reliabilities are somewhat higher than those reported by Brown or Drake (1, 4). The higher reliabilities are perhaps attributable, at least in part, to the use of the reply ' equals ' in cases of doubt.

A factor analysis was carried out by Burt's method of simple summation (3), using successive approximations for the self-correlations. The saturations are appended at the foot of the table. Since, in addition to the factor-saturations so obtained, we have figures for the unreliability of each test, the results of the analysis may be expressed in terms of Burt's four-factor theory (3, p. 103 and refs.). We have in fact factors of four different types : (i) a general factor, (ii) a bipolar factor, (iii) a number of specific factors peculiar to each test, and (iv) factors of error or unreliability.

(i) *The General Factor.* The general factor accounts for about 30 per cent. of the total variance. The question of greatest importance is the psychological nature of this general factor.

As has already been indicated, the common practice is to interpret each factor by a mere scrutiny of the tests into which it enters. This practice has been sufficiently criticized by Burt in this *Journal* and elsewhere; and would certainly be most inappropriate here. Fortunately we have been able to secure results from a number of other tests, musical and non-musical, selected, in accordance with the general plan of the research, as likely to throw light on the problem. Let us begin by comparing the saturations obtained by the six tests of the Seashore battery with their correlations with the various musical and non-musical tests. The result is a suggestive series of figures, set out in Tables II and III.

TABLE II. SATURATIONS FOR FIRST FACTOR AND CORRELATIONS WITH MUSICAL TESTS

Seashore Test	Factor I	Wing	Oregon (*a*)	Oregon (*b*)	Question-naire	Orchestral Instruments
Memory ..	·87	·68	·67	·73	·33	·26
Pitch	·62	·61	·47	·47	·24	·18
Intensity ..	·48	·22	·27	·41	·23	·30
Time	·42	·25	·34	·23	·03	·12
Consonance ..	·39	·44	·46	·43	·36	·14
Rhythm ..	·28	·31	·31	·45	·01	·18
Average ..	·51	·42	·42	·45	·20	·20

The Seashore tests have been rearranged according to their first factor-saturations. If we compare this order with the orders indicated by their correlations with the musical tests, the correspondence is so close as to provide a strong confirmation of the hypothesis that the general factor underlying the Seashore battery is intimately related to musical capacity. If we ignore the test of Consonance (and the justification for this will emerge in the sequel), it will be seen that the questionnaire gives exactly the same order; Wing's tests and the Oregon test (*b*) (nature of alteration) correspond just as closely except for Rhythm; the Oregon test (*a*) (piece preferred) corresponds except for Intensity. On the other hand, knowledge of the instruments of the orchestra test shows little correspondence.

Table III enables us to compare the order of the factor-saturations with that of the correlations obtained with the non-musical tests.

TABLE III. FIRST FACTOR-SATURATIONS AND THE NON-MUSICAL TESTS

Seashore Test	Factor I	Cattell (timed)	Cattell (untimed)	Age	Interest	Cancella-tion
Memory ..	·87	·32	·18	—·05	·34	·15
Pitch	·62	·22	·19	·02	·39	·00
Intensity ..	·48	·13	·25	—·03	·40	·02
Time	·42	·13	·15	·17	·06	—·03
Consonance ..	·39	·03	·00	—·00	·23	·12
Rhythm ..	·28	·17	·19	—·17	—·05	·16
Average ..	·51	·17	·16	·07	·25	·08

The coefficients themselves are now much lower. There appears to be some relation, presumably causal, between the six Seashore saturations and the ratings for ' intelligence '

(especially ' timed ') and for ' interest ' ; but none with age or ' cancellation.' It would seem that speed at the low level measured by the cancellation test is not particularly important for any of the tests, but that speed at the higher levels measured by the Cattell test exercises an appreciable influence on the tests of Memory and Pitch.

The conclusions to be drawn from Tables II and III may be summarized as follows. (i) The general factor in the Seashore battery is related to musical ability on the appreciation side (as measured by the Oregon test) and possibly on the interest and performance side (as measured by the questionnaire). It is similar to the ability measured by the Wing test. (ii) It is, however, also influenced by intelligence, particularly when the test employed takes the form of a speed test.

The next step is to check the analysis by removing the influence of the timed intelligence score from the inter-correlations of the six Seashore tests. This was carried out in the usual way by the aid of Yule's formula for partial correlation. The results are shown in Table IV. A comparison of Tables I and IV shows that, even after ' timed intelligence ' has been removed, the pattern of relations remains much the same as before.

TABLE IV. SECOND-ORDER CORRELATIONS : TIMED-INTELLIGENCE PARTIALLED OUT

Test	Pitch	Intensity	Consonance	Memory	Time	Rhythm
Pitch	—	·36	·27	·36	·23	·10
Intensity ..	·36	—	·18	·18	·26	·08
Consonance ..	·27	·18	—	·43	·24	−·12
Memory ..	·36	·18	·43	—	·19	·39
Time	·23	·26	·24	·19	—	·01
Rhythm ..	·10	·08	−·12	·39	·01	—

(ii) *The Bipolar Factor.* The bipolar factor can be more briefly dismissed.

The saturations are hardly at all affected by partialling out intelligence. On comparing the saturations with the same external characteristics as before (see Table V), only in the case of age and the cancellation test can any appreciable relation be discerned.

TABLE V. THE BIPOLAR FACTOR AND THE NON-MUSICAL TESTS

Seashore Test	Factor II	Cattell (timed)	Cattell (untimed)	Age	Interest	Cancellation
Rhythm ..	−·45	·17	·19	−·17	−·05	·16
Memory ..	−·42	·32	·18	−·05	·34	·15
Consonance ..	·16	·03	·00	·00	·23	·12
Pitch	·18	·22	·19	·02	·39	·00
Time	·26	·13	·15	·17	·06	−·03
Intensity ..	·27	·13	·25	−·03	·40	·02
Average * ..	·29	·17	·16	·07	·25	·08

* Calculated regardless of sign.

The bipolar factor is best interpreted as a factor of classification. It appears to divide the Seashore tests into three subgroups : as will be seen from column 1, the first two tests have high negative saturations, the next two non-significant positive saturations, the last two much higher positive correlations, which are probably quite significant. The first two comprise Rhythm and Memory : here the tests involve a *sequence* of taps in the case of Rhythm, and of tones in the case of Memory. Since the material seems so obviously meaningful, these two tests are commonly selected by musicians (21) as the best in the battery. However, their saturations with the musical factor do

not altogether bear out this verdict: for, although Memory yields the highest figure, Rhythm has the lowest. The second group includes Pitch and Consonance. These are tests involving an immediate comparison of elementary stimuli, in both cases musical tones or combinations of tones. Next to Memory these two tests had the highest correlations with other musical tests (Table II). The third group includes Time and Intensity. Here the task again requires an immediate comparison of elementary stimuli—clicks in the one case and the sounds of a buzzer in the other. This group therefore appears to depend chiefly on stimuli of a non-musical nature. The results might perhaps be tentatively summed up by saying that the bipolar factor appears to contrast tests involving the immediate perception of a change with those involving the immediate memory of a change.

TABLE VI. THE BIPOLAR FACTOR AND THE MUSICAL TESTS

Seashore Test	Factor II	Oregon (*a*)	Oregon (*b*)	Orchestral Instruments	Wing	Question-naire
Rhythm ..	−·45	·31	·45	·18	·31	·01
Memory ..	−·42	·67	·73	·26	·68	·33
Consonance ..	·16	·46	·43	·14	·44	·35
Pitch	·18	·47	·47	·18	·61	·24
Time	·26	·34	·23	·12	·25	·03
Intensity ..	·27	·27	·41	·30	·22	·23
Average ..	·29	·42	·45	·20	·42	·20

(iii) *Error Factors.* The proportion of the variance attributable to unreliability or 'error factors' is rather high. It accounts for more than 20 per cent. of the total variance. Nearly half of this is contributed by the test of Consonance (split-half reliability only 0·34).

The test of Consonance is notoriously difficult to apply, because of the preconceptions on the part of most subjects. The judgments of subjects musically untrained appear to rest on a mixture of vague impressions of 'harmonious combination,' general feelings of pleasantness or unpleasantness, and perhaps other emotional states and associations (6, 7). As Burt and others have pointed out, the test is plainly based on an old-fashioned conception of harmony, which has largely become out of date as a result of the development of modern music (5, 7, 20, 21). In the introspective reports on the test, many of the comments alluded to the influence of context and *Gestalt*, whereby even a rough combination of notes can be accepted as consonant, provided that the discord appears to be satisfactorily resolved either mentally or by the succeeding combination. This was stressed by at least a dozen subjects. A small number of subjects explicitly mentioned the effect of a familiarity with modern harmony on their judgments of relative dissonance. Evidently these students were not strictly obeying the instructions given, namely, to compare the roughness or smoothness of the two combinations themselves.

It may be added that a supplementary analysis was undertaken by examining the partial correlations obtained after intelligence, age, and performance in the cancellation test had been eliminated. However, the elimination of the further variables appeared to leave the general pattern of relations unaffected; and it is therefore unnecessary to report the figures in detail.

(iv) *The Specific Factors.* The factors discussed so far, namely, the general factor, the bipolar factor, and the factors of unreliability, account for only 62 per cent. of the total variance. The remaining 38 per cent. must therefore be explained by specific factors, i.e., by factors which are more or less peculiar to each of the six Seashore tests.

The specific factors may perhaps [1] be interpreted as indicating the amount of the 'atomism' in the tests, so that a comparison with tests of musical ability, and, in particular, with those devised by Wing, should throw considerable light on the extent to which this criticism is justified. With Wing's tests a second bipolar factor, almost as large as the first, was obtained; and in this case both bipolar factors appeared to be fully significant. Hence only about 12 per cent. of the variance

[1] The objection of 'atomism,' urged by Burt and others against the Seashore battery, does not turn solely on the fact that the tests measure highly specific processes; it also implies that they measure only the more elementary constituents (which, of course, tend also to be highly specific). Such processes, so these writers argue, are necessary but not sufficient. Their contention is that "musical appreciation as such is concerned far more with the complex Gestalt-like qualities of musical composition than with the mere discrimination and memory of pitch, intensity, basic rhythms, etc."

in Wing's battery remains attributable to specific factors. Wing's battery, therefore, is not quite as atomistic as Seashore's. But if, as most factorists seem agreed, musical ability is, in Burt's phrase, " highly complex," some degree of specificity or atomism would seem desirable. As has often been stated, " the most economical type of test-battery is one in which all the tests have a high correlation with the general factor, but for the rest have a low correlation with each other."

Final Analysis. One of the aims of the present investigation was to compare the Seashore battery with that of Wing. In all the Wing tests used in the present research the piano is the source of the stimuli ; and passages or chords with some musical significance are employed. A factorial analysis of the scores obtained by the same group of subjects with the seven tests devised by the Wing battery shows that, with his tests, the general factor accounts for as much as 45 per cent. of the total variance ; and, as we shall see in a moment, it is possible, by an appropriate weighting of the Seashore tests, to secure a multiple correlation of 0·72 between the two batteries. It may therefore be concluded that Wing's tests measure much the same kind of ability as Seashore's, but measure it at a higher or at least a different level, namely, that of musical meaning. The results of the entire analysis for both batteries are shown in Table VII.

TABLE VII. A COMPARISON OF THE FACTORIZATION OF THE WING AND SEASHORE TESTS

The figures indicate the proportion (expressed as a percentage) contributed by each factor to the total variance.

Tests	General Musical Ability	First Bipolar	Second Bipolar	Specifics	Unreliability
Wing	45	10	8	12	25
Seashore ..	29	10	0	38	23

TABLE VIII. FINAL ANALYSIS OF THE SEASHORE BATTERY

Tests	Factors				
	G(m)	*G(i)*	*B*	*Sp*	*E*
Memory	·76	·32	·31	·39	·26
Pitch	·59	·22	—·17	·65	·29
Intensity	·46	·13	—·26	·77	·33
Time	·39	·03	—·24	·76	·44
Consonance	·43	·13	—·16	·37	·81
Rhythm..	·27	·17	·52	·62	·41
Percentage of Variance	25	4	10	38	23

The final analysis of the Seashore tests is set out in Table VIII. The first four columns are based on the analysis of the partial correlations in Table IV. *G(m)* denotes that part of the general factor attributable to musical ability as such ; *G(i)*, that part attributable to intelligence ; *B*, the bipolar factor ; *Sp*, the specific factors ; and *E*, the factors of error or unreliability. From these figures a complete picture of the battery can be built up, enabling us to evaluate in precise factorial terms both the claims made on its behalf and the criticisms urged against its shortcomings.

(*b*) *Multiple Correlations with External Criteria.* The preceding results were obtained when each of the six tests of the battery was allowed the same weight. The conclusions drawn, however, show that the value of the several tests as indications of musical ability varies widely from one test to another. The Consonance test, for example, is apparently

devoid both of validity and of reliability, and therefore should probably receive no weight whatever. Indeed, in his *Revised Measures of Musical Talent* (which unfortunately were not obtainable when the above experiments were planned) Seashore himself has dropped it. What we now want to know, before using the test for practical purposes, is (1) how would the correlations be altered if differential weights were employed in combining the scores, and (2) what would be the highest correlation obtainable, if we used the best possible weights ?

As regards weights, the requisite regression coefficients, calculated in the usual way, are shown in Table IX below.

TABLE IX. REGRESSION COEFFICIENTS FOR THE SEASHORE TESTS

Test	Pitch	Intensity	Consonance	Memory	Time	Rhythm
Oregon (*a*) ..	·17	·02	·21	·42	·14	·13
Oregon (*b*) ..	·12	·21	·21	·44	—·02	·25
Wing ..	·39	—·08	·18	·39	·04	·12
Questionnaire	·03	·08	·18	·20	·21	—·07
Interest ..	·20	·33	—·02	·33	—·14	—·25
Intelligence ..	·08	·19	—·09	·08	·07	·11
Age	·04	—·08	—·08	·01	·20	—·19

The multiple correlations are given in Table X, and reveal with greater clarity the relations between the Seashore battery as a whole and the other tests employed. The squares of these correlations indicate the proportion of the variance contributed by the item in question.

TABLE X. MULTIPLE CORRELATIONS AND VARIANCES

Test	Correlation	Variance
Oregon I ..	·75	·55
Oregon II ..	·82	·67
Wing	·72	·62
Questionnaire..	·48	·23
Interest ..	·45	·21
Intelligence ..	·32	·10
Age	—·03	·00

IV. SUMMARY AND CONCLUSIONS

1. Seashore's tests of musical ability were applied to 100 undergraduate and postgraduate students ; and an attempt was made to assess the validity of each component test both by internal criteria (i.e., determining the saturations of each test with the factors entering into all) and by external criteria (i.e., determining the correlation of each test with independent assessments obtained from other objective sources, more especially supplementary tests of musical and non-musical abilities).

2. The inter-correlations were factorized by Burt's method of simple summation ; and, since reliability coefficients can readily be calculated for the separate tests, the results eventually obtained may be conveniently expressed in terms of his ' four-factor type of hypothesis.' They reveal (i) a general factor accounting for 29 per cent. of the variance, (ii) a bipolar factor accounting for 10 per cent., (iii) specific factors accounting for 38 per cent., and (iv) error factors accounting for 23 per cent.

3. The interpretation of the several factors was based partly on the introspective reports secured from the subjects on the mental processes used in performing each of the Seashore tests, but mainly on the external evidence obtained from the supplementary tests. The following conclusions were drawn :—

(i) The saturations for the general factor were first systematically compared with the correlations obtained between each test in the Seashore battery and the tests or assessments of both musical and non-musical characteristics. It was found (*a*) that age and (*b*) musical knowledge have little or no influence on the general factor ; (*c*) that intelligence (especially when assessed by speed of performance) has a small amount of influence ; and (*d*) that performances in the Wing and Oregon tests show the closest agreement. It may therefore be inferred that the general factor is closely related to musical ability and appreciation.

(ii) The bipolar factor is evidently a classification factor. It appears to subdivide the six Seashore tests according to their concrete content, contrasting (*a*) those that depend chiefly upon immediate discrimination with (*b*) those that depend chiefly upon immediate memory.

(iii) The specific factors for Pitch, Intensity, Time, and Rhythm are decidedly large ; those for Memory and Consonance much smaller. It is suggested that the saturations for these factors may be taken as indicating in some degree the ' atomistic ' nature of the battery, i.e., the extent to which it depends upon a number of specific and mutually independent abilities.

(iv) The unreliability of the battery as a whole appears to be rather high. But this is due mainly to the extreme unreliability of the test of Consonance.

4. On these and other grounds it is argued that the common criticisms of the Seashore battery, as expressed by Wing and others—namely, that it is invalid, irrelevant, excessively atomistic, and musically meaningless—are not altogether borne out by the results obtained. A comparison of the Seashore tests with those of Wing suggests that the two batteries measure much the same general musical factor, but at different levels and from rather different aspects. The lower validity of the Seashore battery seems due largely to its greater specificity, and might be easily improved by omitting the test of Consonance and modifying the speed at which the test-items are presented.

5. It is concluded that, in its general nature, the Seashore battery is adequate for its original purpose, namely, to measure the more elementary abilities required for the understanding and appreciation of music ; but that its use will be most effective if the scores are weighted in accordance with the calculated regression coefficients, and if it is used in conjunction with other tests of musical appreciation.

REFERENCES

1. Brown, A. W. (1928). ' The reliability and validity of the Seashore tests of musical ability.' *J. Appl. Psychol.*, XII, 468–76.
2. Burt, C. (1925). *The Measurement of Mental Capacities.* Edinburgh : Oliver and Boyd.
3. Burt, C. (1940). *The Factors of the Mind.* London : University of London Press.
4. Drake, R. M. (1933). ' The reliability and validity of tests of musical talent.' *J. Appl. Psychol.*, XVII, 447–58.
5. Lambert, C. (1948). *Music Ho !* Pelican Books Edition.
6. Pickford, R. W. (1948). ' The borderline of psychology, physics and music.' *Nature*, CLXI, 589.
7. Pickford, R. W. (1949). ' Experiments on the relation of dissonance and context.' *Quart. J. Exp. Psychol.*, I, 57–67 ; III, 107–18.
8. Seashore, C. E. (1919). *The Psychology of Musical Talent.* New York : Silver, Burdett and Co.
9. Seashore, C. E. (1919). *Measures of Musical Talent.* New York : Columbia Phonograph Co.
10. Seashore, C. E. (1938). *The Psychology of Music.* New York : McGraw-Hill.

11. Stanton, H. M. (1935). 'The measurement of musical talent; the Eastman experiment.' *Univ. Iowa Studies*, II.
12. Stanton, H. M., and Koerth, W. (1930). 'Musical capacity measures of adults repeated after musical education.' *Univ. Iowa Studies*, XXXI.
13. Stanton, H. M., and Koerth, W. (1933). 'Musical capacity measures of children repeated after musical training.' *Univ. Iowa Studies*, XLII.
14. Vernon, P. E. (1936). 'Tests in æsthetics.' *ap.* Hamley, H. R. *The Testing of Intelligence.* London: Evans Bros. n.d. 133–41.
15. Vernon, P. E. (1938). *The Assessment of Psychological Qualities by Verbal Methods.* London: H.M. Stationery Office.
16. Whipple, G. M. (1915). *A Manual of Physical and Mental Tests.* Baltimore: Warwick and York.
17. Williams, E. D., Winter, L., and Woods, J. M. (1938). 'Tests of literary appreciation.' *Brit. J. Educ. Psychol.*, VIII, 265-84.
18. Wing, H. D. (1936). *Tests of Musical Ability in Children.* (Unpublished Thesis: University of London Library.)
19. Wing, H. D. (1941). 'A factorial study of musical tests.' *Brit. J. Psychol.*, XXXI, 341–55.
20. Wing, H. D. (1941). *Musical Ability and Appreciation.* (Unpublished Thesis: University of London Library.)
21. Wing, H. D. (1948). 'Tests of musical ability and appreciation.' *Brit. J. Psych. Mon. Sup.*, XXVII.

NIGHT VISUAL CAPACITY AND ITS RELATION TO SURVIVAL IN OPERATIONAL FLYING

By D. D. REID

London School of Hygiene and Tropical Medicine

I. *Problem and Method.* II. *Results.* III. *Summary and Conclusions*

I. PROBLEM AND METHOD

In his Moynihan lectures, Livingston (1944) has described the development and use, in the Royal Air Force during the second world war, of his Rotating Hexagon Test for night visual capacity. Its introduction as a selective method was justified largely on *a priori* grounds, although early results suggested that success in the test was predictive of success in action. Thus the scores for night visual capacity obtained by a group of successful night-fighter pilots were appreciably above the average of flying personnel in general (Livingston, 1948).

The reliability of the test, and its relevance to the performance of tasks of identification in conditions of low illumination, has been fairly thoroughly studied (Bradford Hill and G. O. Williams, 1943). But without clearer indications of the operational importance of testing for night vision, some degree of opposition to the introduction of the Hexagon Test as a method of selecting or excluding men hoping to go forward for flying training was almost inevitable. Moreover, with the increased use of radar aids in aerial gunnery and in landing, ability to see at night became less important during the later stages of the war. Nevertheless, it continues to be relevant both to flying and to other occupations in peace time ; and this would seem to justify the publication of the following results obtained during an attempt at securing an operational validation of the Hexagon Test.

The results of the Hexagon Test of night visual capacity, which was usually applied at an early stage in training, were recorded on each man's medical documents ; these accompanied him throughout his service and were available to the squadron medical officer. In the course of an investigation already reported (Reid, 1947), the records of all the operational crews on active service in Bomber Command of the Royal Air Force were examined and the N.V.C. (Night Visual Capacity) scores were noted, together with the number of bombing sorties completed by the individual by the date of the survey. Then, by producing frequency distributions of N.V.C. scores for men who had completed 0, 1, 2, 3, or more sorties up to the usual tour limit of 30 sorties, it was easy to calculate the mean or average N.V.C. score among groups of men at different stages of their operational tour. If night visual capacity was important in determining the length of survival, i.e., if a selective elimination of those with poor night visual capacity took place by their being shot down early in their tour, then this should be revealed by a progressive rise in the average values of the N.V.C. scores among those who survive to complete most or all of the tour expected of them. If, on the other hand, ability to see at night was quite unimportant in determining survival, the level of the mean N.V.C. scores in groups of increasing lengths of operational experience should remain, within sampling limits, unchanged.

It is clear that night vision is unlikely to be equally important to the wireless operator, whose duty never allows him to use that faculty at all, as it is to the gunners, upon whose speedy appreciation of an impending attack by an enemy night-fighter the safety of the aircraft may well depend. Separate tabulations are thus needed for each type of crew duty.

In view of the many factors of personal skill, mechanical efficiency and all the hazards of weather and war which are involved in a bomber crew's chance of survival, it would be idle to expect any very close relation between an individual's night vision test score and the number of sorties he survived. Consistent rather than spectacular trends must be looked for ; and, even where these trends are clear enough, the comparison of such trends in the various types of crew duty demands the use

TABLE I. MEAN N.V.C. SCORE AT DIFFERENT STAGES OF TOUR

Duty	Number of Sorties completed to date													
	0		1–4		5–8		9–12		13–18		19–30		Total	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean
Pilots	65	18·5	132	19·9	111	18·8	89	19·9	106	19·3	119	20·9	622	19·6
Navigators	94	18·6	123	18·5	100	18·1	91	19·1	102	19·7	112	18·9	622	18·8
Bomb-Aimers	91	18·7	138	18·2	119	17·6	108	19·4	102	19·8	123	19·9	681	18·9
Wireless Operators	114	18·3	192	18·6	148	20·3	111	19·1	143	19·2	146	19·1	854	19·1
Flight Engineers	91	17·5	82	18·3	55	18·0	30	16·9	43	18·7	51	19·9	352	18·2
Mid-upper Gunners	131	18·8	192	18·9	166	19·0	136	19·4	150	19·6	164	20·1	939	19·3
Rear Gunners	149	19·8	195	19·8	164	19·5	134	19·5	166	20·4	178	20·0	986	19·9
Total	735	18·7	1,054	19·0	863	18·9	699	19·3	812	19·7	893	19·9	5,056	19·3

of some statistical method of describing them, and of testing whether apparent differences are larger than could reasonably be due to chance. For these reasons, regression lines have been fitted to the data for each type of crew duty to describe the trend, if any, of the mean night visual capacity score in groups of men of differing operational experience. The significance, in the statistical sense, of these trends and of the differences between them can then be readily assessed by the appropriate analysis of variance technique.

II. RESULTS

The N.V.C. scores were grouped according to the number of bomber operational sorties carried out by the individual up to the time of the survey (0, 1–4, 5–8, etc.) to give series of roughly equal size. The mean N.V.C. scores of the men in each of these 'experience-groups,' divided according to the type of aircrew duty performed, are set out in Table I.

In Table I the totals certainly show no very dramatic rise in mean N.V.C. score in groups of successively longer operational experience. In navigators and wireless operators no clear trend is apparent, but for mid-upper gunners, on the other hand, there is a consistent rise in mean score with each increase in the number of sorties carried out by the date of the survey. To study and compare these differences more precisely, the correlation coefficient between N.V.C. score and operational experience measured in individual sorties was calculated and a straight line fitted to the data to represent the change in mean N.V.C. score with each increase in the number of sorties survived. The regression coefficients (N.V.C. on number of sorties) whose sizes indicate the slopes of this line for each crew duty are set out in Table II.

The Total row in Table II shows that there is a very small, but quite definitely significant, positive relationship between the N.V.C. score at initial examination and the number of sorties survived up to the time of the survey. Further, the apparent progressive rise in mean N.V.C. score with increasing operational experience, such as could have occurred by chance less than once in a thousand times, is best represented by a line whose slope is such that the mean score rises by ·0516 units between groups of men with 1, 2, 3, etc. sorties, to their credit.

TABLE II. RELATION BETWEEN N.V.C. SCORE AND OPERATIONAL EXPERIENCE

Duty	Number	Correlation Coefficient *r*	Regression Coefficient *b*	Critical Ratios *t*
Pilot	622	·0825	·0646	2·0604*
Navigator	622	·0365	·0267	0·9106
Bomb-Aimer	681	·1350	·0939	3·5502**
Wireless Operator	854	·0358	·0268	1·0459
Flight Engineer	352	·1125	·0927	2·1185*
Mid-upper Gunner	939	·0878	·0593	2·6992**
Rear Gunner	986	·0260	·0172	0·8163
Total ..	5,056	·0715	·0516	5·0974**

* Significant : $P < ·05$ ** Highly significant : $P < ·01$

As Table III shows, this trend can be adequately described by a straight line since the residual variability left after fitting it differs insignificantly from the 'error' (in this case the within-array) variance. The same holds for the trends of N.V.C. score in successive groups of increasing operational experience in the specific aircrew duties where the trend is significantly upwards. The regression lines for the four main groups, based on the figures in Table I, are drawn in Fig. 1.

As the tests of significance in Table II suggest, the trends of these straight lines are upward for pilots, bomb-aimers, flight engineers and mid-upper gunners. For wireless operators and navigators there is no such consistent trend : in other words, there is no evidence of a possible selective elimination in action of wireless operators or navigators with below average N.V.C. scores. This, of course, is what might have been expected, since in neither of these aircrew duties is night vision important.

TABLE III. ANALYSIS OF VARIANCE

Source	Sum of Squares	D.F.	Mean Square	F
PILOTS				
Residual	872·3152	29	30·0798	< 1
Regression	175·7040	1	175·7040	4·19**
Between arrays	1,048·0192	30	34·9340	
Within arrays	24,794·9392	591	41·9542	
Total	25,842·9584	621	—	—
NAVIGATORS				
Residual	773·3008	29	26·6655	< 1
Regression	29·9280	1	29·9280	< 1
Between arrays	803·2288	30	26·7743	
Within arrays	22,606·4688	591	36·5592	
Total	23,409·6976	621	—	—
BOMB-AIMERS				
Residual	1,208·8896	29	41·6858	1·36
Regression	391·0592	1	391·0592	12·80**
Between arrays	1,599·9488	30	53·3316	
Within arrays	19,855·4640	650	30·5469	
Total	21,455·4128	680	—	—
WIRELESS OPERATORS				
Residual	937·2624	29	32·3194	< 1
Regression	39·7536	1	39·7536	1·09
Between arrays	977·0160	30	32·5672	
Within arrays	30,031·2272	823	36·4899	
Total	31,008·2432	853	—	—
FLIGHT ENGINEERS				
Residual	1,215·4224	29	41·9111	<1
Regression	205·7056	1	205·7056	4·45**
Between arrays	1,421·1280	30	47·3709	
Within arrays	14,830·3264	321	46·2001	
Total	16,251·4544	351	—	—
MID-UPPER GUNNERS				
Residual	1,010·5504	29	34·8466	1·20
Regression	212·5984	1	212·5984	7·23**
Between arrays	1,223·1488	30	40·7716	
Within arrays	26,344·5280	908	30·3609	
Total	27,567·6768	938	—	—
REAR GUNNERS				
Residual	909·5888	29	31·3651	1·01
Regression	19·8752	1	19·8752	<1
Between arrays	929·4640	30	30·9821	
Within arrays	28,434·4144	955	29·7743	
Total	29,363·8784	985	—	—
TOTAL				
Residual	929·6608	29	32·0573	< 1
Regression	895·0416	1	895·0416	26·00**
Between arrays	1,824·7024	30	60·8234	
Within arrays	173,136·5120	5,025	34·4550	
Total	174,961·2144	5,055	—	—

** Highly significant : P < ·01

Conversely, for mid-upper gunners, the implied importance of night visual capacity agrees with reasonable expectation since their part in 'spotting' enemy fighters might well be decisive for the aircraft's defence by gunnery or evasive manœuvre. Similarly, as Bradford Hill and Williams have shown (1943a), the pilot may require good night vision for landings at night under operational conditions. It is conceivable, too, that the bomb-aimer and flight engineer may from their alternative positions in the front gun cockpit and astradome play a part in the 'spotting' of enemy aircraft which would explain the relevance of night vision to their survival; but the negative findings for rear gunners makes such an explanation less convincing.

As Table II shows, there is no significant upward trend in N.V.C. for the rear gunner category. For this there seem to be two possible explanations. First, it may be that night vision is unimportant in the defensive role played by the rear gunner; but the positive inferences noted for mid-upper gunners render this unlikely. Alternatively, there remains the possibility that the initial selection of men with superior night vision for employment as rear gunners (implied by their higher mean N.V.C. score before commencing the tour) did, in fact, eliminate men whose defective night vision might have proved a danger to the aircraft in which they flew.

TABLE IV. ANALYSIS OF VARIANCE

Source	Sum of Squares	D.F.	Mean Square	F
Deviations from own 'duty' regression	172,824·6992	5,042	34·2770	—
Between 'duty' regressions ..	179·5808	6	29·9301	—
Deviations from common regression	173,004·2800	5,048	34·2718	—
Common regression	895·0416	1	895·0416	26·12**
Within duties	173,899·3216	5,049	34·4423	—
Between duties	1,061·8928	6	176·9823	5·14**
Total	174,961·2144	5,055	—	—

** Highly significant : $P < \cdot 01$

The complete analysis of variance for the data is set out in Table IV. From this table it appears that the differences between the mean score in each crew category are significantly large—a reflection of the policy already noted of selecting those with better results on the Hexagon Test for aircrew duties which, on *a priori* grounds, appeared to demand superior night vision. Then, looking within the various types of duty, part of the differences between the individual scores recorded is clearly accounted for by the regression, common to all duties, between night vision and the number of sorties completed. The variance ratio test shows that the mean square due to this regression is significantly greater than the estimate of random or 'error' variation given by the deviations of individual observations from this common regression line.

The regression line describing the upward trend of mean N.V.C. score with increasing experience may, of course, have a different slope among groups of men with differing crew duties. Indeed, the slope of such trends might be taken to indicate the relative importance of night visual capacity in determining the chance of survival in different duties. A steep slope (such as that shown in Fig. 1 for bomb aimers and flight engineers) would imply that the upward trends in mean N.V.C. score are due to a high rate of elimination of men with inadequate night visual capacity.

It would be tempting to speculate on the possible explanations for the suggestive disparities between the slopes among mid-upper gunners and flight engineers. That such speculation would be premature, however, is evident from the top section of Table IV where the variability between these slopes is shown to lie within the limits of 'error' variation. On the whole, this overall analysis amplifies and confirms the deductions already drawn. It seems evident that in certain types of duty, i.e., heavy bomber pilots, bomb-aimers, flight engineers, and mid-upper gunners, night visual capacity as measured by the Hexagon Test was significantly, if slightly, related to the time of survival in

operational flying. The general course of the upward trend in N.V.C. score seemed to hold for all four types of duty since the differences between them were insignificant, and there was no evidence that this trend was other than linear. On the other hand, although the absence of trend among ' inside ' duties (navigators and wireless operators) agreed with reasonable expectation, it was also absent for rear gunners for whom night vision had been considered to be of prime importance. However, as suggested above, this last feature of the data seems in all probability to be the result of selection. There remains the question—is the general correlation between Hexagon score and duration of survival really the expression of a direct causative relationship between night visual capacity and operational skill?

For the phenomena observed, several explanations may be advanced. It might be that the upward trend in N.V.C. score with increasing operational experience was the result of a relaxation of standards laid down at initial selection. Such a relaxation would presumably result in the more recent recruits to a squadron's strength having a lower average N.V.C. score than their more experienced colleagues. Yet not only is there no record of any such change in policy having taken place, but also such a change would have produced a fairly sudden change in the level of the N.V.C. score rather than the consistent linear trend actually observed. Again, a change such as the complete neglect of N.V.C. testing in selection in general policy would have been expected to have some effect on the trend in *all* types of duty. In fact, as already noted, the trend was conspicuously absent among the ' inside ' members of the bomber crews.

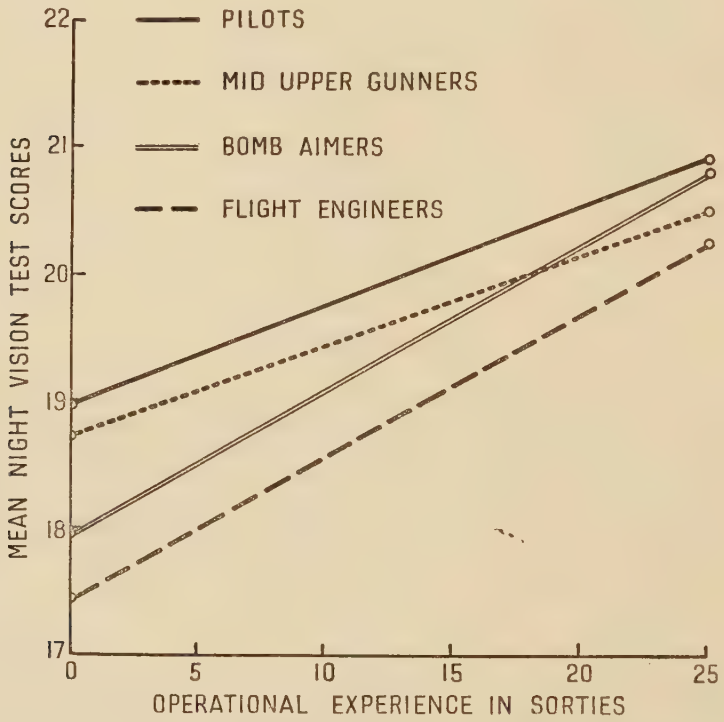

FIG. 1.—Relation between night vision and operational experience.

Yet another explanation would be that the correlation between Hexagon Test score and operational survival depends not on any direct cause and effect link but on some common relation to a third factor, presumably some personal quality or ability which is really effective in determining an individual's chance of completing his operational tour. Intelligence and neurotic predisposition may be two such characteristics. If, for example, intelligence, as measured by the usual paper and pencil tests is positively related both to the results of the Hexagon Test and to the number of sorties likely to be survived, then any selective elimination in action of the less intelligent would result in a simultaneous reduction in the number of men with below average Hexagon Test scores. Similarly, there

is some evidence that neurotic predisposition as assessed by psychiatric interview is related to operational or flying efficiency (Reid, 1942, Davis 1948). There is also evidence which suggests that patients suffering from neurotic illness do rather poorly at the Hexagon Test (Livingston, 1945). It does not follow from this, of course, that these men would have done poorly in the test *before* the onset of their illness, e.g., at initial examination before exposure to operational stress—their performance may be a sign of their acute illness rather than a permanent feature of their native abilities. The possibility that there is some such underlying explanation for the results observed must, however, be investigated.

Some light on the problems involved may be obtained by examining the results of intelligence tests, night vision capacity tests, and psychiatric interviews conducted in the course of another investigation on a large series of pilots under training. Considering the nature of the Hexagon Test, it would not be surprising to find that some of the factors involved in intelligence tests of a conventional kind might have an influence on the results achieved. As it happened, the intelligence tests administered in this particular experiment were divided into a test of form perception (V.C.), a test of verbal intelligence (I.M.A.) and a test of mathematical reasoning (C.M.). As Table V shows, none of these sub-tests shows any significant relation to the results attained by the same individuals in the Hexagon Test. Again, results of a general test of intelligence gave similar negative findings on correlation with Hexagon Test scores (Parry, personal communication). It follows, then, that the correlation of Hexagon score with operational survival does not depend upon any coincident correlation with intelligence in so far as intelligence is measured by the tests used.

TABLE V. CORRELATION OF HEXAGON TEST AND INTELLIGENCE TESTS RESULTS

Test	*r*	*n*	*t*
V.C. ..	·0735	367	1·4077
I.M.A. ..	·0009	367	·0000
C.M. ..	·0419	365	·8091

Note : all coefficients are insignificantly different from zero.

In the other investigation mentioned (Symonds and Williams, 1948) neurotic predisposition was assessed by two psychiatrists who, on the basis of a biographical interview, classified the men examined into four categories of predisposition to the risk of breakdown—nil, slight, moderate, and severe. The N.V.C. scores given by the Hexagon Test for the individuals thus classified were then divided up according to the usual conventions—Excellent (30–32), above average (20–29), average (9–19), below average (3–8) and poor (0–2). Contingency tables, summarizing the relations existing between broad groupings of these categories both of psychiatric assessment and Hexagon Test results, are given separately for each psychiatrist in Table VI.

For both psychiatrists, although the trend in results is suggestive, there is no significant relationship between assessment and N.V.C. score : (for Psychiatrist A, $\chi^2 = 2{\cdot}3877$, $n = 2$, $\cdot50 > P > \cdot30$; for B, $\chi^2 = 1{\cdot}1791$, $n = 2$, $\cdot70 > P > \cdot50$; together, $\chi^2 = 3{\cdot}5668$, $n = 4$, $\cdot50 > P > \cdot30$). In so far as these results give a fair picture of this relation, then there is no reason to suppose that the observed trend in mean N.V.C. score with increasing operational experience is really dependent on the part, if any, played by neurotic predisposition as a factor in survival. The very slight relation appearing in the table may be enough to explain the insignificant rise in N.V.C. score which takes place in all duties with increasing experience.

To sum up, the evidence suggests that the significant upward trend in mean N.V.C. score found in all but one of the ' outside ' types of bomber-crew duty is an expression of some causal relation between the abilities detected by the Hexagon Test and the number of sorties survived, i.e., that night visual capacity, irrespective of any association with either intelligence or neurotic predisposition, is of definite importance in operational flying.

As the low correlation coefficients between N.V.C. score and length of survival suggest, the importance of night vision, relative to all the other factors involved in surviving the many-sided hazards of operational flying, is comparatively small. Indeed, less than 1 per cent. of the total variance

in length of survival is accounted for by differences in night visual capacity as measured by the Hexagon Test. But this is hardly surprising. Quite apart from the many factors involved in operational survival, there is a close interdependence between all members of a crew ; and this combined with the fact that the Hexagon Test was used to eliminate those with more severe deficiencies in night vision, must tend appreciably to dilute the correlation between test result and length of survival. It is indeed a little unexpected that any such correlation could be demonstrated in operation data of this kind.

TABLE VI. RELATION BETWEEN PSYCHIATRIC ASSESSMENT AND N.V.C. SCORE

Psychiatrist A's Assessment	N.V.C. Score: Above Average (8–32)	N.V.C. Score: Below Average (0–7)	Total
Nil	34 (48·57%)	36 (51·43%)	70
Slight	32 (42·10%)	44 (57·90%)	76
Mod. and severe	7 (30·43%)	16 (69·57%)	23
	73	96	169
Psychiatrist B's Assessment	**Above Average (8–32)**	**Below Average (0–7)**	**Total**
Nil	32 (39·02%)	50 (60·98%)	82
Slight	23 (33·82%)	45 (66·18%)	68
Mod. and severe	6 (27·27%)	16 (72·73%)	22
	61	111	172

The operational validation of tests of all sorts in the Services or in industry is expensive in terms of time, labour, and money. Theoretically validation requires the initial testing of all subjects whose subsequent career in the chosen pursuits is then followed up. But in circumstances where new tests are being introduced, there is probably no entirely satisfactory substitute for this approach. Where the need is for a rapid assessment of some personal fact (e.g., a physical measurement or biographical detail), some variant of the method used in this study may be useful, particularly if only current personal records are available.

III. SUMMARY AND CONCLUSIONS

1. In the hope of assessing the importance of night visual capacity as measured by the Livingston Hexagon Test in the determination of operational survival, a statistical analysis was made of the records of over 5,000 men in Bomber Command of the Royal Air Force in 1944. The score achieved by each man tested *before* starting his operational tour was noted, together with the number of sorties survived up to the date of the survey.

2. The mean N.V.C. scores among men who had completed 0, 1, 2, etc., up to 30 sorties were then computed and the presence of an upward trend in these means which might be the result of the selective elimination in action of those with lower N.V.C. scores was tested for significance. Significant upward trends suggestive of a positive correlation between N.V.C. score and the length of operational survival were found among three of the ' outside ' members of a crew (pilots, bomb-aimers, mid-upper gunners) and for flight engineers who had occasional look-out duties, but not for rear gunners and the two ' inside ' members of the crew, viz., navigators and wireless operators.

3. The association noted in these cases does not appear to be the result of any correlation between night visual capacity and intelligence or between night visual capacity and intelligence and neurotic predisposition. There was no demonstrable relation between the scores obtained with the Hexagon and the results of testing of intelligence or neurotic predisposition. The unexpected result in rear gunners may be explained by the rigorous selection standards on the Hexagon Test demanded for rear gunners.

4. The general conclusion is that, whatever its precise mechanism, night visual capacity as measured by the Hexagon Test had a definite, if small, influence on survival in operational flying.[1]

REFERENCES

1. Bradford Hill, A., and Williams, G. O. (1943). 'Statistical studies of night visual capacity as measured by the rotating hexagon test.' *Flying Personnel Research Committee Report*, No. 533.
2. Davis, D. R. (1948). 'Pilot error.' *Air Ministry Air Publication*, 3139A. H.M. Stationery Office.
3. Livingston, P. C. (1944). 'Visual problems of aerial warfare.' *Lancet*, II, 33–8.
4. Livingston, P. C. (1945). 'Practical values and clinical applications of night vision.' *Med. Press*, CCXIII, 4–10.
5. Livingston, P. C. (1948). 'Experiences in night vision in the Royal Air Force.' *Transactions of the Ophthalmological Society of Australia*. VI, 44–58.
6. Reid, D. D. (1942). 'Influence of psychological disorder on efficiency in operational flying.' *Flying Personnel Research Committee Report*, No. 508. H.M. Stationery Office.
7. Reid, D. D. (1947). 'Some measures of the effect of operational stress on bomber crews.' *Air Ministry Publication*, 3139, Chap. XIX. H.M. Stationery Office.
8. Symonds, C. P., and Williams, D. J. (1948). 'Psychological disorders in flying personnel.' *Air Ministry Publication*, 3139, Chap. XVI, 205–27. H.M. Stationery Office.

[1] I wish to thank the Director-General of Medical Services of the Royal Air Force, Air Marshal P. C. Livingston, both for his help and for his permission to publish this paper. In its preparation I have had the advice of the Royal Air Force Consultant in Medical Statistics, Professor A. Bradford Hill. To Mrs. K. M. Bull, Miss B. M. Miller, and Miss O. M. Penfold, I am indebted for computing and secretarial assistance, and to Mrs. M. G. Young for drawing the diagrams.

PRIMARY FACTORS OF PERSONALITY

By H. A. REYBURN and M. J. RAATH
The University of Cape Town

I. *The Experiment.* II. *The Centroid Analysis.* III. *The Orthogonal Analysis.* IV. *The Oblique Analysis:* (a) *Orthogonal Projection;* (b) *Parallel Projection.* V. *The Factors.* VI. *The Interrelations of the Factors.* VII. *Summary.*

I. THE EXPERIMENT

The purpose of the experiment discussed in this paper was to reconsider and extend the results of factorial analyses of personality qualities already made by us (9, 10, 13). A larger range of qualities has been dealt with, the choice being governed partly by clinical experience and partly by analyses published by others, particularly Cattell (5).

There appear to be three main methods of dealing with the axes which a centroid analysis yields.

First, the axes, as they stand, may be taken to represent intelligible factors. The resulting general factor and subordinate bipolar factors are then regarded as concordant with the demands of logical definition (1). Against this must be set an important and perhaps decisive objection. The factors thus reached are relative to the battery of tests or variables, and thus lack objectivity (11, p. 397, and 12, p. 55).

Secondly, the axes may be rotated to find simple structure. For reasons given elsewhere we are not satisfied with this solution (12, 14); and in practice it does not seem to give the objectivity claimed. Indeed, the minimum conditions of simple structure are not always rigidly insisted upon, and the criteria for the supposed unique result turn out, in fact, to be a matter of degree (16, pp. 335 and 355).

Thirdly, factors may be identified positively and axes drawn through guiding points in the light of all available knowledge (12, pp. 59f.; cf. 15, pp. 287f.). The criterion is then the intelligibility of the analysis as a whole and its concordance with other results. Certainty and objectivity are not to be expected at first. Factors which will remain objective in all relevant circumstances must be sought gradually and by a method of approximation. It is our hope that this further paper may contribute to this process.

For the present enquiry two groups of students (37 men and 46 women) acted as observers; all had some acquaintance with psychology, and most were advanced students. Each was asked to select two subjects, approximately of his own age, and thoroughly well known to the observer. The total number amounted to 160, viz., 78 men and 82 women. About half were persons outside the University, and the sexes were fairly evenly balanced. In a few cases two observers chose the same subject: the correlations between their ratings were calculated as an indication of the reliability, and averaged ·806. The observers were first asked to sketch the main personality features of the two subjects chosen; these were recorded under pseudonyms, and were remarkable for their freedom and candour. A questionnaire was then filled up, which required a number of qualities to be rated on a 5-point scale.

In Cattell's analysis bipolar qualities were frequently used: it was apparently assumed that psychologically, if not logically, the opposites were contradictories, not mere contraries. We were unwilling to make this assumption. Accordingly, the negative was taken to be the absence or extreme infrequency of the quality. It was left to the analysis itself to show if contraries could be treated as contradictories.

In administering the questionnaire the two groups of observers were taken separately. Each observer was given a typed copy of the questions and one of us went over the items with the whole group, giving further explanations. Each item was answered by everyone before the next question was considered ; and stress was laid on the need of keeping the answers independent. There was one difference between the two groups. Item 26 deals with the level of ideal which the subject demands. The first group answered the question in a general way ; the second divided it into two, reporting separately on the standard which the subject exacts from himself and on that which he exacts from others.

The ratings for each pair of qualities were correlated, first for each of the two groups of 80 subjects, and then for the full 160. The results served as a check on the reliability. In addition the answers were scrutinized in the light of the sketches already made.

II. THE CENTROID ANALYSIS

In making an analysis by simple summation or the so-called centroid method there is no infallible criterion to determine the number of factors to be extracted, and in any given experiment more may be present than the accuracy of the data allows one to extract. Weight must be given to the probability at each stage that the residuals can be explained as chance variations. This criterion was adopted by Reyburn and Taylor in analysing some of Webb's data (13, p. 157) ; but theoretical and practical considerations have since induced us to think that the analysis should be carried slightly beyond the stage thus indicated. In general it is desirable to bring the average residual below the probable error of a zero correlation ; and attention should be paid to the rate at which the average contribution to the variance diminishes.

In the present experiment, after five factors had been extracted, the average residual was ·051 ; the P.E. of a zero correlation for 160 cases would be $\pm$ ·053. One more factor may perhaps be extracted. When this is done, the average residual falls to ·046. The average amount of variance extracted by each successive factor is (in percentages) as follows : Factor I, 14·8 ; II, 10·5 ; III, 10·9 ; IV, 5·4 ; V, 3·2 ; VI, 2·2. The saturations obtained for the several traits are shown in Table I. These may be called F_0.

The test of a successful rotation is the intelligibility of the factors obtained. Each should be a unitary function, and not a mere aggregate of accidentally associated qualities. It is difficult to rotate successfully if nothing is known beforehand of any of the factors. If, however, it is known that certain factors are present, the task is simplified, and it may be possible to establish new factors. In the method we adopt it is thus necessary to begin with some hypothesis regarding at least a number of factors, and the first and immediate test of its adequacy is the intelligibility of the set of factors as a whole. The *n*th factor must be as intelligible a unity as the first.

No single experiment, however, can finally establish the objectivity of any factors. Objectivity becomes finally manifest only when we can assume a set of factors in a variety of contexts and find on each occasion that the remaining factors also prove significant. The ultimate test is the coherence and intelligibility of the factors through the widest range of experimental material. In searching for factors which are objective in this sense, one must not expect to find them with high precision at the outset. If some variable is known to have a heavy load on a factor, it can be used as a guide to determine the best position of the axis. But the guide will not be perfect. The final position for the axes can only be approached gradually and in a series of approximations. In making an analysis, therefore, it is important to use past experience as well as present data.

The first decision to be made is whether the rotated axes are to be orthogonal or oblique. There are obvious advantages in favour of the former : the system is simpler, and the independence of the factors is a considerable asset. With the latter the angles between the axes are a part of the total picture ; and whether they are resolved into factors of the second order or not, they add to the complexity of the analysis, and thus seem to offend against the principle of parsimony.

Orthogonal factors, however, are apt to be abstract and difficult to identify. Often they seem to consist in what is left of certain variables when what is due to other factors has been set aside. And although this presents no mathematical difficulty, it often presents psychological difficulties. Further, the order in which the factors are established is of the greatest importance. The possible load of the variables on any factor is limited by what has already been taken out of them. Thus the first factor tends to absorb the maximum loads of the variables, and the last one is only allowed to have what is left after all the correlations with the other factors have been deducted. These objections may not be fatal, but they are formidable. And they induced us to make a double analysis, using both orthogonal and oblique axes.

III. THE ORTHOGONAL ANALYSIS

We begin by fixing provisionally points through which the axes may be passed. To determine these points we made a survey of the data. The correlations were classified according to the significance which could be attached to them.

TABLE I. CENTROID FACTORS

No.	Trait	I	II	III	IV	V	VI	h^2
1.	Avoids Problems	·390	·278	−·575	−·200	−·038	·189	·638
2.	Overbold	·336	−·252	·323	·104	−·061	·069	·300
3.	Unsystematic	·373	−·084	−·380	−·289	·230	·128	·443
4.	Spontaneous	·445	−·529	−·067	·111	·082	·101	·512
5.	Emotional	·331	−·211	−·450	·441	−·113	−·182	·597
6.	Emotionally variable	·328	−·240	−·093	·027	·110	−·092	·195
7.	Emotionally uncontrolled	·632	−·039	−·071	·261	−·122	·005	·490
8.	Relaxes easily	·270	−·148	−·224	−·428	·228	·008	·380
9.	Looks on bright side	·211	−·518	−·030	·000	−·065	−·093	·326
10.	Avoids strong emotions	−·503	·323	·086	−·090	−·091	−·118	·395
11.	Easily distressed	·142	−·048	−·263	·457	−·188	−·065	·341
12.	Prone to sad thoughts	·064	·415	−·259	·397	·057	−·058	·408
13.	Fits of depression	·206	·337	−·295	·303	·158	−·273	·434
14.	Avoids tests	·377	·302	−·387	−·198	−·151	·000	·445
15.	Feels inferior	−·050	·420	−·456	·243	−·032	·391	·601
16.	Shy	−·341	·417	−·396	−·018	−·065	−·120	·467
17.	Assertive	·327	−·118	·680	·070	·144	·000	·609
18.	Submissive	−·309	−·144	−·570	−·125	−·060	−·155	·485
19.	Subjective outlook	·329	·185	·362	·198	−·100	−·114	·336
20.	Fixed in outlook	−·395	·157	·209	·120	−·064	·306	·337
21.	Suggestible	·138	−·071	−·560	−·149	−·153	·201	·424
22.	Contra-suggestible	·252	·326	·380	·115	·175	·060	·362
23.	Incompliant	·311	·338	·432	·020	·275	−·253	·538
24.	Dominating	·480	·119	·472	·125	·022	·261	·551
25.	Self-critical	−·378	−·180	−·222	·158	·001	·172	·279
26*a*.	Realistic : Self	·443	·163	−·155	−·255	·666	·062	·759
26*b*.	Realistic : Others	·561	·208	−·179	−·367	·370	−·178	·693
27.	Aggressive	·527	·311	·335	·194	·072	·021	·530
28.	Withdrawn	−·339	·553	·098	·054	·000	−·072	·439
29.	Self-esteeming	·255	−·182	·598	−·136	−·055	−·175	·508
30.	Conceited	·510	·224	·300	−·202	−·314	·157	·564
31.	Self-reliant	−·374	−·173	·488	·203	·226	·003	·501
32.	Reliable	−·556	−·040	·264	·293	·142	·059	·490
33.	Persevering	−·490	−·159	·315	·274	·226	−·267	·563
34.	Cheerful	·360	−·700	·115	−·130	·038	−·053	·654
35.	Irritable	·455	·349	·036	·271	−·038	−·156	·429
36.	Excitable	·565	−·284	−·229	·250	−·150	·031	·538
37.	Easily rattled	·485	·111	−·258	·105	−·216	−·065	·377
38.	Suspicious	·214	·448	·235	·184	·089	·228	·396
39.	Kind	−·329	−·469	−·202	·106	−·101	−·104	·402
40.	Sensitive to others	−·478	−·308	−·225	·273	·064	·087	·461
41.	Touchy	·122	·310	−·258	·544	−·147	−·074	·500
42.	Buoyant	·001	−·616	·029	−·268	−·091	−·003	·460
43.	Blames others	·585	·313	·050	·042	−·236	−·038	·501
44.	Unbalanced	·426	·282	−·386	−·132	−·168	·091	·465
45.	Mixes badly	−·372	·585	−·166	−·029	·058	−·133	·530

In the first class were placed all which, in a sample of 160, could be expected to be significant at or below the ·001 level ; in the second class those at or below the ·01 level ; in the third, those at or below the ·02 level ; in the fourth, those at or below the ·03 level ; in the fifth, those at or below the ·05 level ; in the sixth, those at or below the ·10 level ; in the seventh class, those at or below the

·20 level ; and all the rest were regarded as devoid of significance. In the analysis attention was given almost exclusively to the first two classes.

The First Factor. One of the factors isolated from Webb's study on character and intelligence was named ' cleverness ' by Garnett (7) and ' surgency ' by Cattell (3 and 4).

In 1933 and 1936 Cattell indicated the close interrelationship between surgency and a number of other qualities. The relation was set forth as a bipolar one. This is shown in the following list which indicates the more intimate connections. *Positive Traits :* cheerful ; natural ; sociable ; humorous ; adaptable ; gregarious ; quick ; hasty. *Negative Traits :* gloomy ; formal ; unsociable ; earnest ; conservative ; exclusive ; slow ; introspective.

In 1939 Reyburn and Taylor found that the following qualities had significant factor loads on an axis placed to measure surgency-desurgency. *Positive Loads :* fondness for large social gatherings ; sense of humour ; cheerfulness ; corporate spirit ; bodily activity in pursuit of pleasure. *Negative Loads :* fits of depression ; conscientiousness. In 1941 Reyburn and Taylor dealt with the factor from the desurgency end, and found the following loads. *Positive Loads :* Indulges in self-pity ; loses head in excitement ; worries ; is easily hurt ; prefers working with others. *Negative Loads :* Takes prominent part in social affairs. In 1943 Reyburn and Taylor obtained the following results for surgency : *Positive Loads :* carefree ; happy-go-lucky ; unconcerned about others' opinions ; likes excitement ; unconcerned about future ; also talkative. *Negative Loads :* over-conscientious ; worries ; easily distressed ; thinks before acting ; dislikes being interrupted.

In 1944 Cattell made a fresh analysis of the ' Personality Sphere ' (4), and obtained a classification of traits for the bipolar factor (re-named surgency versus agitated, melancholic desurgency), which will be found in his book (5). A third and slightly different arrangement was presented in 1947 (6).

A consideration of these various analyses shows that there are several aspects to the activity which is being treated as a functional unity. On the one hand, there is an emotional aspect ; on the other, a more directly conative one, which may be called responsiveness. In our experiment, variable No. 4 (spontaneous) corresponds closely to the latter aspect ; the former one seems adequately covered by variables Nos. 9, 12, 13, 34, and 42. Their inter-correlations merit examination ; and are shown in Table II.

TABLE II. CORRELATIONS BETWEEN VARIABLES CLASSED AS RESPONSIVENESS

No.	Variable	4	9	12	13	34
9.	Looks on bright side ..	·423				
12.	Prone to sad thoughts..	−·156	−·280			
13.	Has fits of depression ..	−·038	−·237	·492		
34.	Cheerful..	·539	·510	−·418	−·399	
42.	Buoyant..	·349	·389	−·434	−·315	·595

The variables fall into two groups : Nos. 9, 34, and 42 present the surgent aspect ; Nos. 12 and 13, the desurgent aspect. The bond between the two sections is fairly strong, but weaker than within themselves. Moreover, they show a different relation to No. 4. Its correlation with the group consisting of Nos. 9, 34, and 42 is ·536 ; whereas with the two desurgent variables, Nos. 12 and 13, it is only −·123 : i.e., the two emotional aspects, though related, do not move together as a functional unity. Spontaneity carries cheerfulness with it, but does not exclude depression.

These considerations suggest that it is inadvisable to attempt, at least at the beginning, to locate a bipolar surgency-desurgency factor, and that it is better to use spontaneity as a guide to determine the positive aspect of the main factor in question. A trial of two arrangements, one in which spontaneity alone was used as a guide, and another in which it was combined with the centroid of variables Nos. 9, 34, and 42, showed little difference in the results. Accordingly the simpler form was used, and an axis was passed through No. 4, the factor obtained being provisionally named Spontaneity.

The Second Factor. The second factor represents Stability or integration. To reach it, an axis has been passed through the centroid of variables Nos. 5 reflected (not emotional) ; 13 reflected (free from fits of depression) ; 20 (fixed in outlook) ; 32 (reliable) and 37 reflected (not easily rattled).

This factor is in line with two of Cattell's factors, viz., *C* and *G* (5, pp. 317, 326, 480, and 489) and the main qualities attributed by him to them are reasonably well covered by the relevant relationships of our second factor as set forth in Tables V and VII below.

Although the positive aspect of surgency is largely accounted for by the first factor, the negative aspect remains almost intact. A desurgent factor, named '*d*' (mental depression), was described by the Guilfords in 1939 (8). But such a factor, when tried out on our data, although intelligible in itself, does not combine well with the factors that follow it, and in particular it makes the final two difficult to interpret. The various traits in the questionnaire which have a neurotic tinge are not reducible to a single factor ; nor, if desurgency is abstracted from them, is the remainder readily intelligible. Moreover, neurotic deviations from the normal presuppose some positive principle of integration.

The Third Factor. The remaining factors may be described more briefly. The third factor is derived from Webb's conception of *w* (17). In 1939 Reyburn and Taylor, examining Webb's material, altered the conception chiefly by giving more weight to steadiness, continuity, and perseverance than to action from principle or purpose (9, p. 162). In 1943 they found the same factor appearing, although less clearly, in some of the Guilfords' material (13). A factor of this kind has also been postulated by other writers who have not set it in a factorial context. Variable No. 33 is taken as a guide to it here, and Persistence is suggested as a suitable name.

The Fourth Factor. One of Cattell's factors, *E*, is described as 'dominance (hypomania) versus submissiveness' (5, p. 321) ; and a similar quality has been recognized by others. It is represented fairly well in our list by variable No. 24 (dominating). But we have one alternative in No. 17 (assertive). As can be seen from the figures the two traits—dominating and assertive—agree in a good many points. But there are nearly as many significant failures to agree. Probably dominance is not a simple or ultimate unity, but involves assertiveness along with other elements. Eventually it was decided to pass an axis through No. 17. The factor of Assertiveness thus obtained resembles that of self-confidence as described by several other authors (18).

The Fifth and Sixth Factors. In the analysis of Webb's material (9, p. 162) Reyburn and Taylor found a factor which they named charity. In the list of qualities used in the present experiment there is a variable, viz., No. 39 (Kind), which might be expected to prove a guide to this factor if it is present in this analysis. Consequently an axis was passed through No. 39. The result, however, was unsatisfactory.

The fifth factor offers a choice, for it can be regarded from either end. On the one side, there are variables No. 11 (easily distressed), No. 36 (excitable), No. 5 (emotional), No. 41 (touchy), and on the other, No. 26 *a* and *b* (low standard of behaviour), No. 8 (relaxes), and No. 3 (unsystematic). The factor is bipolar, with sensitiveness, in perhaps more than one form or manifestation, at the one end, and toughness, callousness, and indifference at the other. The contrast of tender and tough presents the situation fairly well ; and it seemed advisable to make Sensitiveness the positive end.

The sixth factor affords a measure of a feeling of Inferiority, and gives a clear, coherent picture.

The transformation matrix (T_1) which gives effect to these ideas is shown in Table III : the factor loadings resulting from this transformation may be called F_1.

TABLE III. ORTHOGONAL TRANSFORMATION MATRIX

·623	−·515	−·283	·511	−·052	·054
−·739	−·420	−·218	·314	−·044	·359
−·094	·571	·123	·785	·134	−·122
·155	−·105	·560	·008	·543	·597
·114	·161	·300	·058	−·813	·455
·141	·441	−·673	−·140	·150	·538

IV. THE OBLIQUE ANALYSIS

When oblique axes are used, there are two ways of dealing with the situation. On the one hand, each axis may be treated as if it were the first, no allowance being made for the correlation of the variables with the other factors ; on the other, each factor may be treated as the last, and its loads calculated only after the effect due to the correlations of the variables with the other factors has been taken away. We may consider these two processes in turn.

(*a*) *First Method.* Using the same guiding points as before, we may pass an axis through each of them without reference to the position of the other axes. In doing this we have a choice of two

starting points, viz., the co-ordinates given by the centroid analysis itself, or those given by the orthogonal rotation. The use of the latter simplifies the arithmetic, and the transformation matrix required (T_2) is shown in Table IV below.

TABLE IV. OBLIQUE TRANSFORMATION MATRIX

1·000	−·447	−·249	·326	·242	−·271
0	·894	·519	·366	−·338	−·348
0	0	·817	·128	·342	−·348
0	0	0	·862	−·331	−·396
0	0	0	0	·774	·180
0	0	0	0	0	·704

The transformation matrix is triangular (*see* Burt, 1, p. 306); the factors obtained are oblique; and the cosines of the angles between them are given by the inner products of its columns, as follows:

TABLE V. CORRELATIONS BETWEEN FACTORS

Trait	1	2	3	4	5	6
1. Spontaneity.. ..	1·000					
2. Stability	−·447	1·000				
3. Persistence	−·249	·576	1·000			
4. Assertiveness ..	·326	·181	·213	1·000		
5. Sensitivity	·242	−·411	·044	−·287	1·000	
6. Inferiority	−·271	−·190	−·398	−·602	·203	1·000

(*b*) *Second Method.* If the axes obtained by the method just used are regarded as primary axes in Thurstone's sense of the term, the loadings obtained are the correlations of the variables with those axes; and F_2 will give the 'structure' of these factors. To find the 'pattern' for the same factors we must post-multiply Table II by the transpose of the inverse of T_2 (Table IV). This transformation may be termed T_3.

The difference between the factor loadings obtained by the two transformations can be expressed as follows. With six factors there are six hyperplanes each containing the axes of five factors and the axes themselves can be regarded as the six lines in which these hyperplanes intersect one another. Since the factors are all oblique, each axis is not orthogonal to the hyperplane which does not contain it, but forms an angle with the normal to that hyperplane. In Thurstone's procedure, where guidance is sought from the conception of simple structure, the hyperplanes are first determined, and from these the normals and primary factors are derived. The orthogonal projection of the variables gives the structure of the factors, and the parallel projection gives the pattern—in which the simple structure is supposed to be found. The three sets—the hyperplanes, the normals to the hyperplanes or reference vectors as Thurstone calls them, and the primary factors—are related in such a way that if any one set is determined, the others are also determined. Thurstone, as has been said above, determines the hyperplanes first. But it is possible to begin elsewhere, and we have begun with the primary factors. As we have seen, the post-multiplication of F_1 by T_2 gives the factorial structure, i.e., the loads obtained by orthogonal projection on to the primary axes (the correlations between the variables and the factors). The post-multiplication of F_1 and T_3 gives the parallel projection of the variables on to the primary axes. There is, however, an advantage in making this parallel projection not on to the primary axes themselves but on to the reference vectors. To do this the loadings given by $F_1 T_3$ would have to be post-multiplied by the cosines of the angles between the primary axes and their corresponding reference vectors.

Each of the foregoing factorial arrangements has its own value. If simple structure is sought, parallel projection, of course, is essential. But on the whole, when knowledge is incomplete and factors are only in the process of discovery, preference, we think, should be given to the orthogonal projection of the variables on to the primary axes, i.e., to the factor loadings described as F_2. Our reasons may be briefly summarized as follows:

(1) In parallel projection, although the loads on the axes can be accurately calculated if the whole system is known, the factors themselves, on the whole, are difficult to identify and recognize

in the concrete. As in the case of the later factors in an orthogonal system, each factor is, as it were, what is left when several other things have been taken away ; and generally it is not easy to realize what this residuum is : whereas with immediate orthogonal projection on to the primary axes, the psychological meaning of the factors can usually be grasped more readily.

(2) The successful use of parallel projection requires that the factor analysis has been completed and all the factors found. If, for example, the centroid analysis has extracted only five factors, when six are really present, and we attempt to find a factor-pattern on five oblique factors, the whole scheme goes awry ; for even if good guiding points for the five axes can be obtained, the direction of projection on to them is not known. The hyperplane, parallel to which we project, contains only $(n-2)$ instead of $(n-1)$ dimensions. Of course, with direct orthogonal projection on to the axes representing the primary factors, there is also error ; but the error is usually not so great. In this case what is missing is the guidance which would have been obtained from an additional centroid factor. With parallel projection, however, there is no guarantee that the missing factor is smaller than the rest ; and moreover, the axis of each of the others is slightly out of its true position.

(3) It may well be that we cannot place all the primary factors correctly, even if we can extract the right number by the centroid process. In such a case parallel projection is unworkable ; whereas the method of orthogonal projection on to the primary axes, when these axes are positively determined, may be used for any axes that are known. If an adequate centroid analysis has been made, a study of a single factor can be carried out with considerable success, even if the other factors remain undetermined.

V. THE FACTORS

The six factors which we have found, together with their relations to the original variables, may be set forth as follows, the figures in brackets indicating the correlation between the factor and the relevant variable.

Factor I—Spontaneity. With a high degree of this quality the subject is full and free in his reactions ; he is brimming over, full of life, and is spontaneous and responsive in most situations. On the other hand with a low degree of it he is restricted, inhibited or apathetic and unable to let himself go. The spontaneous person tends to be (*a*) cheerful (·708) ; (*b*) a good mixer (·665) ; (*c*) free from the tendency to withdraw (·631) ; (*d*) excitable (·609) ; (*e*) likes stimulating and exciting things (·601) ; (*f*) is not shy (·511) ; (*g*) is responsive to the bright side (·496) ; (*h*) expresses emotions readily (·457) ; (*i*) is easily stirred up (·435), and (*j*) tends to recover his cheerfulness easily (·400).

Factor II—Stability. With a high degree of this quality the subject is well integrated, balanced, and mature. The stable person is (*a*) not generally emotional (·627) ; (*b*) nor excitable (·576) ; (*c*) nor easily flurried (·566) ; (*d*) restrains expression of emotions (·558) ; (*e*) is usually self-reliant (·544) ; (*f*) reliable (·544) and (*g*) faces up to problems (·528), and to tests (·488) ; (*h*) requires a high standard of behaviour (·487), and (*i*) has fixed ways of looking at things (·476) ; (*j*) is slow to blame (·460) ; (*k*) free from fits of depression (·456) ; (*l*) persevering and persistent (·432), and (*m*) not irritable (·431).

Factor III—Persistence. With a high degree of this quality the subject tends to carry through tasks which he has begun. He is not easily put off by difficulties, nor does he abandon activities quickly and readily when the first novelty has worn off. The persistent person (*a*) tends to face up to the problems of life (·708), and to tests (·591) ; (*b*) is usually balanced (·632) ; (*c*) self-reliant (·628), and (*d*) reliable (·612) ; (*e*) not usually suggestible (·483), nor (*f*) conceited (·479) ; (*g*) slow to blame (·470), and (*h*) systematic (·468) ; (*i*) not easily rattled (·453), and (*j*) requires a high standard of behaviour (·445).

Factor IV—Assertiveness. With a high degree of this quality the subject is quick and ready to maintain his point of view. He is not inclined to let things pass with which he does not agree. The assertive person (*a*) is free from false modesty (·633), and (*b*) not commonly submissive (·627) ; (*c*) tends to dominate and lead (·609), and (*d*) is not shy (·565) ; (*e*) is indifferent to the views of others (·508) ; (*f*) is aggressive (·497), and (*g*) free from feelings of inferiority (·466) ; (*h*) is not suggestible (·461) but (*i*) contra-suggestible (·431) ; (*j*) ambitious (·456) ; (*k*) has his own way of looking at things (·425), and (*l*) faces up to the problems of life (·405).

Factor V—Sensitiveness. With a high degree of this quality the subject is easily touched and responds to the actions and emotions of other people. The contrast of tender versus tough indicates the two ends of the scale with fair adequacy. The sensitive person (*a*) is generally emotional and easily stirred up (·607) ; (*b*) has a high standard of behaviour for himself (·548), and (*c*) expects it of others (·413) ; (*d*) is distressed by the difficulties of others (·534) ; (*e*) is touchy (·479), and (*f*) is excitable (·407) ; (*g*) compliant and conciliatory (·319) ; (*h*) does not relax (·309) ; (*i*) is naturally kind (·308).

Factor VI—Inferiority. With a high degree of this quality the subject suffers from distressing feelings of inferiority. These may be covered and even compensated for by aggressive action ; but they underlie much that he does. The person who has a sense of inferiority (*a*) has a low opinion of

himself and his abilities (·596) ; (*b*) lacks cheerfulness (·540) ; (*c*) avoids the problems of life (·498) ; (*d*) is slow to assert himself (·470) ; (*e*) is prone to sad thoughts (·467) ; (*f*) touchy (·451), and (*g*) does not easily regain cheerfulness (·433) ; (*h*) is shy (·418) ; (*i*) often unbalanced and immature (·364), and (*j*) is a poor mixer (·360).

VI. THE INTERRELATIONS OF THE FACTORS

The interrelations of the factors have been given in Table V above. But since, to reach this result, the original data have undergone a number of computational processes, it is desirable to obtain some check of the reliability and objectivity of the analysis as a whole. To this end use has been made of fresh material. A considerable time after the original experiment was carried out, it was repeated with different observers, reporting on 62 new subjects. Then, each factor being taken in turn, weights were provisionally allotted to the qualities associated with them as set forth in section V above, and scores were calculated for each subject. These, when inter-correlated, gave the following results :

TABLE VI. INTERRELATIONS OF FACTORS

Trait	1	2	3	4	5	6
1. Spontaneity	1·000					
2. Stability	−·181	1·000				
3. Persistence	−·087	·587	1·000			
4. Assertiveness	·117	·182	·220	1·000		
5. Sensitivity	·039	−·120	−·036	−·174	1·000	
6. Inferiority	−·490	−·240	−·264	−·295	·120	1·000

The agreement of this table with Table V is probably as good as can be expected with the data and the method adopted. All significant figures have the same signs ; most of the numerical values agree fairly well, the differences reaching the 5 per cent. level of significance only in two cases. Taking the two tables together the following generalizations may be ventured.

Spontaneity tends to be associated with assertiveness and to reduce stability and inferiority. Stability tends to be associated with persistence and possibly with assertiveness, and to reduce spontaneity, sensitiveness, and inferiority. Assertiveness tends to be associated with spontaneity, persistence, and possibly with stability, and to reduce inferiority and sensitiveness. Sensitiveness is possibly associated with inferiority, and tends to reduce stability and assertiveness. Inferiority tends to reduce assertiveness, persistence, stability, and spontaneity.

When a system of oblique axes is used, it is desirable to ask whether the inter-correlations of the factors can be accounted for by a secondary factor or set of such factors. But in the present instance at least three secondary factors would be required ; and we have so far discovered no convincing positions into which their axes could be rotated. All the many positions which they might take up are equally arbitrary ; and, in the light of existing knowledge, it seems inadvisable to lay any weight on these factors.

The ultimate value of factors derived by analysis lies in the light which they throw on the individual. They provide a framework within which the study of him may be set, and obviously the first step will be to assign him a score in respect of each factor. This, however, does not exhaust the situation. There is what may be called a normal relation between the factors themselves, discovered in the first instance statistically by reference to the total population examined. But there is no ground for believing that this statistical or normal relation must hold in detail for every individual. The interrelations of the factors in each individual are also a part of the true picture of him, and those relations merit examination.

The most obvious method would be the direct one. The individual may be measured a considerable number of times in respect of each of the factors, and the correlations resulting from those measurements may be compared with the normal values : that is to say, something analogous to what Burt has called P-technique (1, pp. 169–209 ; 2, pp. 178–203) may be applied, not to discover factors, but to investigate the relations between them in the individual. The process would be laborious ; and, although it seems the most fundamental method, it is not always practicable.

Without applying it generally, it may perhaps be possible by considering individual results to gain some idea where further analysis on these or other lines is desirable. When the score of an individual on a factor deviates significantly from the mean, we can ask whether this deviation is accompanied by any other unexpected deviation in his scores on other factors. In the case of orthogonal factors, the normal exception is the mean itself; and hence in this case the question which we have to ask is whether more factors than one deviate from the mean at the same time. If two or more such factors do so, this may be due to accidental variations in the separate factors; but it may also be due to an unusual connexion between the factors in the individual. In the case of oblique factors, when one deviates from the mean, the most probable position of the other is not the mean itself, but a distance from it given by a regression equation. If the other factors do not have the values which these considerations lead one to expect, a qualitative analysis of the individual is generally indicated. If the factors obtained by analysis are used in this way their value may be greatly enhanced. But a fuller discussion of the point must be reserved for another occasion.

VII. SUMMARY

1. The purpose of the experiment discussed in this paper was to modify and extend the analysis of personality qualities already made by the method used by the authors. Eighty observers each gave a personality sketch of two subjects, rating them also on 45 or 46 qualities. The resulting correlations yielded six centroid factors.

2. Rotation of the centroid axes was undertaken both into orthogonal and into oblique positions, the latter being dealt with in two ways, viz., to give structure on 'primary factors' and also on 'reference vectors.' The reasons for selecting the positions into which the axes were rotated have been indicated above. In the end preference was given to the method which projects orthogonally on to oblique primary axes.

3. It is claimed that the resulting factors and their inter-correlations may provide a framework within which the study of the individual may be set.

REFERENCES

1. Burt, Cyril (1940). *The Factors of the Mind.* University of London Press.
2. Burt, Cyril (1940). 'The factorial study of temperamental traits.' *Brit. J. Psychol., Stat. Sect.*, I, 178–203.
3. Cattell, R. B. (1933). 'Temperament tests.' *Brit. J. Psychol.*, XXIII, 308–29.
4. Cattell, R. B. (1944). 'Interpretation of the twelve primary personality factors.' *Character and Personality*, XIII, 55–91.
5. Cattell, R. B. (1946). *Description and Measurement of Personality.* Harrap and Co.
6. Cattell, R. B. (1947). 'Confirmation and clarification of primary factors.' *Psychometrika*, XII, 197–220.
7. Garnett, J. C. M. (1918). 'General ability, cleverness and purpose.' *Brit. J. Psychol.*, IX, 345.
8. Guilford, J. P., and Guilford, R. B. (1939). 'Personality factors D, R, T and A.' *J. Abnorm. Soc. Psychol.*, XXXIV, 21–36.
9. Reyburn, H. A., and Taylor, J. G. (1939). 'Some factors of personality.' *Brit. J. Psychol., Gen. Sec.*, XXX, 151–65.
10. Reyburn, H. A., and Taylor, J. G. (1941). 'Factors in introversion and extraversion.' *Ib.*, 147-9.
11. Reyburn, H. A., and Taylor, J. G. (1941). 'Factorial analysis and school subjects: a Criticism.' *Trans. Roy. Soc. S.A.*, XXVIII, 168–85.
12. Reyburn, H. A., and Taylor, J. G. (1943). 'On the interpretation of common factors: a criticism and a statement.' *Psychometrika*, VIII, 53–64.
13. Reyburn, H. A., and Taylor, J. G. (1943). 'Some factors of temperament: a re-examination.' *Psychometrika*, VIII, 91–104.
14. Reyburn, H. A., and Raath, M. J. (1949). 'Simple structure: a critical examination.' *Brit. J. Psychol., Stat. Sect.*, II, 125–33.
15. Thomson, Godfrey H. (1948). *The Factorial Analysis of Human Ability.* University of London Press.
16. Thurstone, L. L. (1945). *Multiple-Factor Analysis.* University of Chicago Press.
17. Webb, E. (1915). 'Character and intelligence.' *Brit. J. Psychol. Mon. Suppl.* No. 3.
18. Wolfle, D. (1942). 'Factor analysis in the study of personality.' *J. Abnorm. Soc. Psychol.*, XXXVII, 393–7.

THE SIGNIFICANCE OF FACTOR LOADINGS

LAWLEY'S TEST EXAMINED BY ARTIFICIAL SAMPLES

By STEN HENRYSSON

Department of Psychology and Education, University of Uppsala, Sweden

I. *Introduction.* II. *The Construction of Artificial Samples.* III. *Conclusion.*

I. INTRODUCTION

One of the weak points in factor analysis is the lack of satisfactory significance tests [1] for deciding whether the residuals left in the correlation matrix after extracting a certain number of factors can be explained by sampling errors or not. A step towards a solution of this problem was taken by Lawley (1940, 1941, 1943). Starting (1940, p. 65) from Wishart's joint probability distribution of the sample variances and covariances,[2] he finds a likelihood function, which is treated as a function of the population parameters (the factor loadings) and for this he endeavours to find a practicable solution.

His analysis starts with a preliminary estimate of the number, m, of common factors, and loadings for these factors are then found by an iterative process, beginning with a set of trial values. A method, suitable for large samples, is suggested for testing the assumption initially made in regard to the precise number of factors required. From his likelihood function he deduces that, if the assumption is correct, then the quantity

$$N \sum_{i<j} \left\{ \frac{(c_{ij} - a_{ij})^2}{s_i^2 s_j^2} \right\} , (i, j = 1, 2, 3, \ldots, n) \qquad (1)$$

(where N is the number of observations, c_{ij} are the covariances calculated from the loadings, a_{ij} are the given covariances, and s_i^2 are the error variances) will be distributed as χ^2 with p degrees of freedom, where

$$p = \tfrac{1}{2}\{(n - m)^2 - (n - m)\} . \qquad (2)$$

" If a significantly high value of χ^2 is obtained, this will indicate that we must reject the hypothesis and assume the existence of more than m factors " (1941, p. 184). This test of significance is applicable only if the above method of estimation is used.

The object of the following paper is to examine Lawley's test of significance by means of artificial samples. The test is intended to be used on large samples ; and our object is to find out if it works on samples of 200 observations. This will be done by ascertaining how the values of χ^2, which we get for our different samples, are distributed.

II. THE CONSTRUCTION OF ARTIFICIAL SAMPLES

Wold (1948) gives 25,000 numbers taken at random from a normally distributed population ($M = 0$, $\sigma = 1$). The numbers are given in ten uncorrelated columns, $x_1, x_2, \ldots, x_{10}$. We construct nine variables containing one general factor according to the factor pattern

$$z_i = x_1 + x_{i+1} \quad (i = 1, 2, \ldots, 9) \qquad (3)$$

[1] See, however, the paper by Professor M. S. Bartlett in the last number of this *Journal* (III, 1950, pp. 77–85) on ' Tests of Significance in Factor Analysis,' especially sect. II.

[2] The method is not restricted to correlation coefficients.

by adding to the number in column (1) first, the number in column (2), secondly, the number in column (3), and so on until we reach column (10). The value in column (1) is thus common to all the nine variables. We divide the first 24,000 random numbers into twelve parts with 200 rows in each, and thus get *twelve* samples of nine variables each containing 200 observations :

$$z_{iN} = x_{1N} + x_{(i+1)N} \quad (N = 1, 2, \ldots, 200) \qquad (4)$$

In each sample of 200 observations the variances and covariances of the variates z_i are given by the formula

$$a_{ij} = \frac{1}{199}\left[\Sigma z_i z_j - \frac{1}{N}(\Sigma z_i)(\Sigma z_j)\right] \quad (i, j = 1, 2, \ldots, 9), \qquad (5)$$

where Σ here denotes summation over the sample. The calculations may be simplified by using equation (3). Thus we have

$$a_{ij} = \frac{1}{199}\left[\Sigma x_1^2 + \Sigma x_1 x_i + \Sigma x_1 x_j + \Sigma x_i x_j - \frac{1}{200}(\Sigma x_1 + \Sigma x_i)(\Sigma x_1 + \Sigma x_j)\right]. \qquad (6)$$

The sums in the equation above (which are sums over 200 numbers) may be obtained by adding together sums over four successive groups of 50 numbers. These latter have been calculated by Wold (1948) and Azorin and Wold (1950). For the first sample of 200 the matrix of variances and covariances [1] was found to be as follows :

TABLE I. MATRIX OF VARIANCES AND COVARIANCES

Variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)
(i)	3·894	1·819	2·192	1·957	1·878	1·987	1·845	2·091	1·959
(ii)	1·819	3·833	1·829	1·688	1·757	1·794	1·503	1·769	1·562
(iii)	2·192	1·829	4·136	2·291	1·779	2·066	1·793	2·406	1·896
(iv)	1·957	1·688	2·291	3·948	2·069	1·968	1·787	1·957	1·835
(v)	1·878	1·757	1·779	2·069	3·885	1·867	1·716	1·864	2·012
(vi)	1·987	1·794	2·066	1·968	1·867	3·739	1·917	2·003	1·815
(vii)	1·845	1·503	1·793	1·787	1·716	1·917	3·358	1·608	1·835
(viii)	2·091	1·769	2·406	1·957	1·864	2·003	1·608	3·993	1·907
(ix)	1·959	1·562	1·896	1·835	2·012	1·815	1·835	1·907	3·417
Loadings	1·438	1·230	1·495	1·419	1·354	1·402	1·264	1·428	1·343

Assuming one general factor, the loadings estimated by the maximum likelihood method were as shown in the last line of the table. The value of χ^2 required to test the hypothesis is found to be 21·8, which, with 27 degrees of freedom, is not significant. The probability P that such a value would be exceeded by chance is 0·74. For the whole set of twelve samples the values of χ^2 and the corresponding values of P are as given below.

TABLE II. VALUES OBTAINED FOR CHI-SQUARED

Sample	1	2	3	4	5	6	7	8	9	10	11	12
χ^2	21·8	38·6	24·2	19·2	28·1	17·8	31·6	31·9	28·2	22·1	25·8	15·5
P	0·74	0·07	0·62	0·86	0·41	0·92	0·25	0·24	0·40	0·73	0·53	0·96

[1] When calculating the variances and covariances by equation (6), we divide by 100 instead of 199 to make this step easier. This operation has no influence on the final values of χ^2.

III. CONCLUSION

The results obtained are in good agreement with theory. The values of P should conform with a rectangular distribution ; and the expected range is 0·85 (Cramér, 1946, p. 372), a figure which is close to the observed range (0·96 — 0·07) of 0·89. If all twelve samples are combined, a total value for χ^2 of 304·8 is reached. This, with $12 \times 27 = 324$ degrees of freedom, corresponds to a probability P of 0·77. From the test we should therefore conclude that all the 9×9 covariance matrices can be expressed in terms of one significant common factor only, with nine specific factors ; and this conclusion accords with what in these artificial cases we know to be the fact.[1]

REFERENCES

1. Azorin, F., and Wold, H. (1950). ' Product sums and modulus sums of H. Wold's normal deviates.' *Trabajos de Estadistica*, I, 5–29.
2. Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton : Princeton University Press.
3. Lawley, D. N. (1940). ' The estimation of factor loadings by the method of maximum likelihood.' *Proc. Roy. Soc. Edin.*, LX, 64–82.
4. Lawley, D. N. (1941). ' Further investigations in factor estimation.' *Proc. Roy. Soc. Edin.*, LXI, 176–85.
5. Lawley, D. N. (1943). ' The application of the maximum likelihood method to factor analysis,' *Brit. J. Psychol.*, XXXIII, 172–5.
6. Thomson, G. H. (1948). *The Factorial Analysis of Human Ability*, Ch. xxi. London : University of London Press.
7. Wold, H. (1948). ' Random normal deviates,' *Tracts for Computers*, XXV.

APPENDIX

By CYRIL BURT

Maximum Likelihood Loadings as Obtained from Single-Factor Correlation Tables

Mr. Sten Henrysson, in his suggestive paper, has applied Lawley's test of significance to a set of factor loadings calculated by his first ' method of maximum likelihood ' from an artificial covariance matrix containing one common factor only. In this country factor analysis has been most frequently carried out, not with covariances, but with correlations. Accordingly, as many readers of this *Journal* may be interested in that form of application, I have ventured to convert Mr. Henrysson's matrix of covariances [2] into a matrix of correlations. The coefficients so obtained are set out in Table III, with the corresponding factor-saturations added at the foot.

A word or two is perhaps required to indicate how British psychologists have commonly approached such a problem. In early discussions on the statistical validation of mental tests, interest centred almost entirely on the ' general ability ' (or ' intelligence '), which seemed to dominate all others. Accordingly in 1915 I ventured to put forward a simplified procedure for analysing correlation tables which contained no more than a single significant factor, i.e., tables of the type which Mr. Henrysson has here constructed. In such cases it seemed clear that, for the purposes of the psychologist, Pearson's method of obtaining ' the line of closest fit ' by the method of least squares needed modification. When we are dealing not with physical but with psychological measurements (the results of different intelligence tests, for example), we are almost bound to assume that a unique or

[1] I should like to thank Professor Wold for his help and encouragement in this research, which was carried out in the Department of Statistics in the University of Uppsala, and Dr. D. N. Lawley and the editors, Professor Thomson and Professor Burt, for their assistance in preparing my manuscript for publication.

[2] Mr. Henrysson states (footnote p. 160) " when calculating the variances and covariances, we divide by 100 instead of by 199 " (or 200). Hence to obtain variances and covariances as calculated by the more usual procedure, his figures would have to be halved or (more strictly) to be multiplied by 100/199.

specific factor is present in each test. The larger the specific factor, the more it tends to reduce what (in a somewhat broad sense) was originally called the ' reliability ' or ' precision ' of the test. From this it was argued that, in accordance with the usual least squares procedure, the correlation table should in theory first be ' prepared ' by dividing each standardized variable throughout by the square root of its unreliability. And the argument was later extended to tables in which several common factors might be involved. In the case of an *ideal* single-factor matrix it was easy to show that this theoretical correction led to the same figures for the factor-saturations as would be obtained by applying Pearson's procedure direct to the correlation table with *reduced self-correlations* in the principal diagonal instead of unity ; and further that, in such a case, this procedure itself could be further simplified by substituting the formula for simple summation for the more elaborate calculations required by Pearson's method of principal axes. It was therefore suggested that, even in dealing with an *empirical* table where the observed coefficients did not differ appreciably from a perfect hierarchical arrangement, the same shortened procedure might legitimately be adopted.[1]

As applied to the problem of fitting a table of correlations the method of simple summation was identical with that already in use for fitting a contingency table (1, p. 64, eq. 1) ; and it was natural to suggest that the discrepancies or ' residuals ' should be tested for significance in the same familiar fashion, namely, by the chi-squared test (1, p. 64, eq. 3). On the same grounds as before, however, it appeared desirable to retain the correction for specificity for the residuals before testing their significance in this way. This resulted in a formula for residual correlations which was identical in form with the Pearson-Yule formula for partial correlation : (cf. 2, p. 57). The similarity, of course, is not accidental : precisely the same principles are involved in the two arguments.

The chi-squared test furnishes a significance test for the system of residuals taken as a whole, and was thus used chiefly with a bipolar analysis. To locate significant residuals occurring either in isolation or in clusters, it is more instructive to test each separately by an appropriate probable or standard error ; this method was accordingly used in group factor analyses. Since each augmented residual (or ' specific correlation,' as it was called) could be regarded as a partial correlation with the general factor eliminated, it seemed legitimate (with large samples) to test it by taking the ordinary standard error for a partial correlation of zero, that is, by taking $1/\sqrt{N}$, where N denotes the number of persons tested : (1, p. 352, eq. 15). This implies that the augmented residuals can be treated as conforming approximately to a normal distribution with a standard deviation of $1/\sqrt{N}$; and, on that assumption, the sum of their squares, multiplied by N, would conform to the chi-squared distribution. Thus, without any very rigorous proof, it was assumed, and the assumption seemed borne out in practice, that the two procedures would yield much the same results.

The object of this note is to show that, in cases like that selected by Mr. Henrysson, the results obtained by this simple procedure [2] are identical with those obtained by the more elaborate method of calculation which he has adopted in his paper.

[1] This procedure was first briefly outlined in a paper, submitted to the Psychological Subsection of the British Association, on ' General and Specific Factors Underlying the Primary Emotions ' : (the summary, as printed in *Brit. Ass. Ann. Rep.*, 1915, pp. 694–6, gives the figures so obtained, and discusses the concrete results, but does not describe the technicalities of the method in any detail ; a fuller account was given in my L.C.C. Report on ' The Relations between Ability in Different Subjects in the School Curriculum,' reprinted, with some abridgment, in *The Distribution and Relations of Educational Abilities*, 1917).

It is perhaps necessary to remind those who refer to these earlier papers that the use of terms has since slightly changed. In conformity with Galton's terminology, the phrases ' specific abilities,' ' specific factors,' and ' specific correlations ' were originally employed to refer to what would now be called ' group factors.' The phrase ' specific factor,' however, was later adopted by Spearman and his pupils to denote a factor peculiar to a single test, i.e., what is now sometimes called a ' unique factor ' ; and ' reliability ' was used to mean, not ' precision,' but ' self-consistency.' Hence what was originally called ' correction for unreliability ' was later called ' correction for specificity.'

[2] I am tempted to suggest that the short and simple algebraic proof, put forward to support the older method, may perhaps be easier for the student to follow than the more rigorous approach adopted by Lawley. The assumptions made in the text imply that the matrix to be analysed is not (in the usual notation) R, the matrix of observed correlations with unity in the diagonal, but $R_u^{-\frac{1}{2}}(R-R_u)R_u^{-\frac{1}{2}} = R_u^{-\frac{1}{2}}R_cR_u^{-\frac{1}{2}} = R_a$ say, where R_u denotes the diagonal matrix of specific factor variances, and $R_c = R - R_u$, the correlation matrix with reduced self-correlations. This formulation leads to equations of the usual type, namely,

$$f_a'(R_a - v_aI) = 0 \text{ and } |R_a - v_aI| = 0 ; \qquad \text{(i)}$$

but the formula now relates to correlations corrected for ' unreliability.' And the procedure requires, not Spearman's correction formula $\frac{r_{ij}}{\sqrt{r_{ii}\, r_{jj}}}$, but the complementary formula, $\frac{r_{ij}}{\sqrt{1-r_{ii}}\sqrt{1-r_{jj}}}$, where r_{ii} and r_{jj} denote ' reliability ' interpreted as above (viz., ' reliability ' so far as the tests i and j depend on common factors only).

TABLE III. THE 'OBSERVED' CORRELATIONS

Variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)
(i)	(·5314)	·4708	·5462	·4991	·4828	·5207	·5102	·5303	·5370
(ii)	·4708	(·3950)	·4594	·4339	·4553	·4739	·4189	·4522	·4316
(iii)	·5462	·4594	(·5392)	·5670	·4438	·5254	·4811	·5920	·5043
(iv)	·4991	·4339	·5670	(·5102)	·5283	·5122	·4908	·4929	·4996
(v)	·4828	·4553	·4438	·5283	(·4736)	·4899	·4751	·4733	·5522
(vi)	·5207	·4739	·5254	·5122	·4899	(·5277)	·5410	·5184	·5078
(vii)	·5102	·4189	·4811	·4908	·4751	·5410	(·4749)	·4391	·5417
(viii)	·5303	·4522	·5920	·4929	·4733	·5184	·4391	(·5085)	·5163
(ix)	·5370	·4316	·5043	·4996	·5522	·5078	·5417	·5163	(·5288)
Simple Summation	·7302	·6285	·7335	·7140	·6887	·7268	·6886	·7123	·7273
Weighted Summation	·7290	·6285	·7343	·7143	·6882	·7264	·6891	·7131	·7272
Loadings: from Correlations	1·438	1·230	1·493	1·419	1·356	1·404	1·263	1·425	1·344
from Covariances	1·438	1·230	1·495	1·419	1·354	1·402	1·264	1·428	1·343

In the present case the working is as follows. Starting with the ordinary table of correlations (Table III), saturations have first been computed by 'simple summation' (last line but three in the table). These have then been taken as a basis for obtaining the improved values furnished by 'weighted summation' (last line but two). It may be noted that, although the matrix is not perfectly hierarchical, the use of weighting introduces hardly any change in the figures. Each saturation has then been multiplied by the square root of the variance as given in Mr. Henrysson's table. Thus for variable (i) we obtain $\cdot 7290 \times \sqrt{3 \cdot 894} = 1 \cdot 438$, and so on. It will be seen that (except for trifling discrepancies) *the loadings thus finally reached are identical with those obtained by Mr. Henrysson from his matrix of variances and covariances* (see last two lines of Table III).

Next, taking the improved values for the saturations, a 'hierarchical' matrix of correlations has been constructed in the usual way to fit the observed correlations. The differences

Now, if R_c is a matrix of rank one, the ordinary method of expanding the characteristic determinant leads at once to

$$0 = |R_c - v_c I| = v_c^n - (r_{11} + r_{22} + \ldots + r_{nn})\, v_c^{n-1},$$

which gives $v_c = 0$ or $v_c = \Sigma r_{ii} = \Sigma r_{ig}^2 = f_c' f_c$; and similarly $|R_a - v_a I|$ leads to $v_a = 0$ or $v_a = f_a' f_a$. But $f_a f_a' = R_a = R_u^{-\frac{1}{2}} R_c R_u^{-\frac{1}{2}} = R_u^{-\frac{1}{2}} f_c f_c' R_u^{-\frac{1}{2}}$. Hence we may take $f_a' = f_c' R_u^{-\frac{1}{2}}$ and $v_a = f_c' R_u^{-1} f_c$. From this it is easy to show that, under these conditions, the results of factorizing R_c will satisfy equations (i). Substituting the values thus found for f_a' and v_a, we have

$$f_a' R_a = f_c' R_u^{-\frac{1}{2}} \cdot R_u^{-\frac{1}{2}} R_c R_u^{-\frac{1}{2}} = f_c' R_u^{-1} f_c \cdot f_c' R_u^{-\frac{1}{2}}$$

$$= f_c' R_u^{-1} f_c \cdot f_a' = v_a f_a' :$$

that is, $f_a' (R_a - v_a I) = 0$, as required. Thus, when the reduced correlation matrix is hierarchical, there is no need to carry out the elaborate calculations entailed by equations (i): the observed correlation matrix can be factorized just as it stands, except that reduced self-correlations must first be substituted for unity. Since in such a case $r_{ii} = r_{ig}^2$, the expression for the residuals becomes $\dfrac{r_{ij} - r_{ig} r_{jg}}{\sqrt{1 - r_{ig}^2}\sqrt{1 - r_{jg}^2}}$. (Note that, if R_c involves *several* common factors, $F_c' R_u^{-1} F_c$ is *not* a diagonal matrix, and thus does not satisfy the 'orthogonality condition' in the form imposed by Lawley.)

between these theoretical figures and the observed figures in Table III (i.e., the first ' residuals ') have then been divided by the product of the square roots of the two relevant specific factor variances. Thus

$$r_{12.g} = \frac{\cdot 4708 - \cdot 7290 \times \cdot 6285}{\sqrt{1 - \cdot 7290^2}\ \sqrt{1 - \cdot 6285^2}} = \cdot 0238.$$

These ' corrected residuals ' are shown in Table IV.

TABLE IV. THE ' SPECIFIC ' CORRELATIONS $\left(r_{ij.g} = \frac{r_{ij} - r_{ig} r_{jg}}{\sqrt{1 - r_{ig}^2}\ \sqrt{1 - r_{jg}^2}}\right)$

Variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)
(i)	—	·0238	·0232	−·0449	−·0376	−·0183	−·0157	·0215	·0149
(ii)	·0238	—	−·0042	−·0276	·0405	·0327	−·0254	·0070	−·0476
(iii)	·0232	−·0042	—	·0892	·1249	−·0172	−·0510	·1430	·0637
(iv)	−·0449	−·0276	·0892	—	·0727	−·0135	·0028	−·0341	·0431
(v)	−·0376	·0405	·1249	·0727	—	−·0194	·0017	−·0342	·1044
(vi)	−·0183	·0327	−·0172	−·0135	−·0194	—	·0812	·0009	−·0428
(vii)	−·0157	−·0254	−·0510	·0028	·0017	·0812	—	−·1035	·0816
(viii)	·0215	·0070	·1430	−·0341	−·0342	·0009	−·1035	—	−·0054
(ix)	·0149	−·0476	·0637	·0431	·1044	−·0428	·0816	−·0054	—

With 200 cases the standard error for a partial correlation of zero may be taken as $1/\sqrt{200} = 0{\cdot}0707$. Of the 36 figures in Table IV, only one is more than twice the standard error, which is in agreement with Mr. Henrysson's value for P (0·74). Chi-squared can be calculated by squaring each of the ' corrected residuals,' adding the squares, and multiplying by $N = 200$. The figure so obtained is 22·0 as compared with Mr. Henrysson's figure of 21·8. Thus chi-squared for the residual correlations is practically identical with chi-squared for the residual covariances, as was after all to be expected. (Here, and in the foregoing figures for the loadings, the slight discrepancies between his figures and my own are no larger than might be anticipated from a difference in the extent to which the iterative calculations have been carried. Indeed, had we dispensed with iterations altogether and merely taken the figures given by simple summation, the result would have been virtually the same.)

Tables III and IV above correspond respectively with Tables XVIII and XX and Tables XXII and XXIII of my early Report (2, 1917, pp. 52 and 57 and pp. 61 and 62). The factor-saturations there given were derived by inserting reduced self-correlations in place of the ordinary reliability coefficients or self-correlations of unity.

The use of reduced self-correlations was strongly criticized at the time by both Pearson and Spearman, and has since again been questioned in the last issue of this *Journal* by Professor Thomson.[1] But it may, I think, be claimed that the agreement between the results reached by this substitution and those furnished by the first ' method of maximum likelihood ' can be plausibly adduced as one further argument in favour of reduced self-correlations, at any rate in such cases as the present.

The variations in the saturation coefficients are themselves not without interest. It will be remembered that, in constructing his covariance table, Mr. Henrysson gave an equal weight to the general and to the specific factor in every case. Hence the expected value for each of the nine saturations will be $1/\sqrt{2} = 0{\cdot}707$. The standard deviation of the observed values about this true value is 0·032. With 200 cases the standard error of an ordinary correlation of 0·707 would be ±0·035. At first sight the similarity between the two figures might seem to support the suggestion that a rough estimate of the standard error for the saturation coefficient can be obtained by regarding it as an ordinary coefficient of correlation. However, figures obtained from dice-throwing experi-

[1] In *Human Ability* (1950) Spearman repeats his criticism of " the filling of the vacant diagonal places by stopgaps," which he thinks " quite arbitrary " (pp. 29–30). For Thomson's criticisms see *Brit. J. Psychol., Stat. Sect.*, III, 1950, p. 126, and refs.

ments by Mr. H. F. Weston and myself some time ago indicate that such an estimate is decidedly too low. The approximate formula I myself have suggested (*Mental and Scholastic Tests*, p. 139) here gives a value of $\pm 0{\cdot}058$. But further investigations along these lines would be highly instructive.[1]

REFERENCES

1. Yule, U. (1912). *Introduction to the Theory of Statistics.* London : Griffin.
2. Burt, C. (1917). *Report on the Distribution and Relations of Educational Abilities.* London : P. S. King.

[1] In a study of significance tests Mr. Weston has calculated artificial tables using *different* proportions for the general factor in each test, and including in some cases a second bipolar factor. Here too the figures obtained by applying the least squares procedure (' weighted summation ') to the correlation table, with reduced self-correlations in the diagonal, prove to be very similar to the figures reached by the more laborious procedure described by Lawley and Emmett (this *Journal*, II, pp. 90–7 ; cf. Burt, *ibid.*, pp. 106, 113). With the chi-squared tests he finds the ' correction for specificity ' rather more generous than the z-transformation. It would be instructive if Mr. Henrysson could give detailed figures for the loadings in *all* his twelve tables. At the same time it should be remembered (as Dr. Lawley points out in a note he has been good enough to send me) that the errors of estimation in the nine loadings would not be uncorrelated. He adds that according to his own formula the standard error should be about $\pm 0{\cdot}063$.

THE FACTORIAL ANALYSIS OF QUALITATIVE DATA

By CYRIL BURT

Psychological Department, University College, London

I. *The Importance of Qualitative Data in Psychology.* II. *Alternative Statistical Techniques.* III. *The Treatment of Multiple Determinates.* IV. *A Factorial Analysis of Physical Attributes.* V. *Summary and Conclusions.*

I. THE IMPORTANCE OF QUALITATIVE DATA IN PSYCHOLOGY

The Form of the Data. In many investigations within the field of individual differences, the available data are expressed, not as quantitative measurements stating magnitude or degree, but in terms of classes or attributes which are essentially qualitative. Statistical psychologists, and particularly those who have worked with standardized tests and employed factorial methods, are often accused of ignoring the qualitative aspects of their problems. As a rule, their critics seem to assume that, because such observations are not recorded in the form of quantitative assessments, they are no longer amenable to quantitative treatment, and can therefore have no use or interest for the statistical investigator. That, however, is a patent fallacy. If they are to be tabulated, such observations, it is true, must be set down in what the schoolboy calls the ' nought-and-one ' style, where each ' one ' will signify that yet another individual belongs to the class named or possesses the attribute specified, and ' nought ' will signify that he does not. But, when that has been done, it is a simple matter to count the ones and compare their sum with that of the ones and noughts added together : in this way we can readily summarize our observations in the form of a table of frequencies or probabilities ; and these can manifestly be subjected to statistical treatment : (cf. 3, 4, 5).

In psychology the oldest and most familiar instance of this procedure is furnished by the marks given for the Binet-Simon tests. Here each examinee is in effect awarded a measurement of ' one ' for every test he passes and ' nought ' for every failure. A similar device has long been adopted by teachers in marking the simpler type of examination paper : each child's total score is obtained by counting every correct answer as ' one,' and adding up the total. So-called personality-tests —the Rorschach, the personal inventory, the biographical questionnaire, and enquiries about interests and attitudes, for example—are frequently scored in this way.

As actually observed, no doubt, most mental traits (probably as a result of their highly complex origin) take the form of continuously graded variables. But the patterns, the combinations, and the interactions between such traits (or between the statistical factors to which it is convenient to reduce them) often have a qualitative character, which cannot itself be expressed as a difference of degree. Moreover, individual psychology is not concerned merely with the description of individuals considered in isolation. It has to take into account the causal influences which affect each individual and his observable behaviour ; and many of these causal influences are not themselves reducible to quantitative variations. For example, in my early studies of delinquency, it was often essential to take up points which lead to a classification rather than a measurement. Is the child illegitimate ? Is he an only child ? Is he living with his parents ? Has he suffered from this or that disease ? What is the lad's present occupation ? These and many other questions cannot be answered by a figure on a rating scale, but only by a categorical ' yes ' or ' no,' or (where there are several alternatives) by a ' this and therefore not that or any other.'

Nor are qualitative data of this nature confined solely to psychology. They occur in many related fields. In physical anthropology, for example, the investigator has constantly to deal with what statistical writers have termed ' attributes ' as contrasted with ' variables.' Pearson's well-known paper on ' the correlation of characters not quantitatively measurable ' (3) was prompted by the need to deal with qualitative traits like hair-colour and eye-colour. And, as a glance at the earlier chapters of Yule's text-book (6) will show, some of the stock illustrations of such procedures are still drawn from the anthropological field. More recently in biological research, and particularly in work on experimental genetics, the investigator has been concerned quite as often with the assignment of individuals to distinct classes as with the measurement of variable traits.

The distinction between ' measurement-data ' and ' frequency-data ' and between ' continuous distributions ' and ' discrete distributions ' has long been recognized in statistics. And the growing realization that mathematics is not restricted to the study of quantity has made it increasingly possible to discuss qualitative characteristics in the same rigorous way, and often along the same lines, as quantitative. On the wider issues I may perhaps refer the more advanced student to the contributions of French statisticians, who have dealt most fully with the problem.[1]

Now we must, I think, agree with our critics that, in studying individual differences whether in physical or in mental characteristics, factorial investigations have hitherto concentrated chiefly on quantitative measurements. It would, however, be plainly advantageous if all kinds of characteristics, qualitative as well as quantitative, could be included in such studies. And the chief purpose of this paper will be to indicate how factorial methods may still be applied, even when much of the data is expressed in a qualitative form.

The Nature of the Factors. As with quantitative variables, so with qualitative attributes the factors thus derived will be essentially principles of classification. Indeed, the factor analysis of attributes is related, much more obviously than the factor analysis of variables, to the modes of classification employed in the simpler sciences.

In botany, for example, the trained observer inspects in order the parts of the plant he is seeking to identify—its root, its stem, its leaf, its flower, and its fruit ; and knows that each of these parts may exhibit one of a number of alternative forms or qualities. The veining of the leaf may be parallel or netted ; the colour of the flower may be white, red, yellow, blue ; and so on. Even where some of the features have numbers attached, e.g., 3, 4, or 5 petals, these numbers still specify, not differences in degree, but differences in kind. The diagnostic combinations of such attributes correspond to the factors.

The chief divergence between problems of classification in the simpler sciences and in the anthropological sciences is that in the former a far higher degree of certainty attaches to the inferences to be drawn. Thus *all* tulips have parallel veins ; nearly all hawthorns have flowers that are white or red ; and, with few exceptions, all clover-leaves are three-lobed. In chemistry exceptions are rarer still : all metals of the alkali class are univalent ; all occur in the first group of the periodic table ; all act as powerful bases ; all are strongly electro-positive, and the strength of this tendency almost invariably increases with increasing atomic weight (here, we may note in passing, the classification turns in part on differences in quantitative variables as well as on qualitative attributes). On the other hand, we do not find that *all* children who fail in the Binet tests for Age VIII also fail in those for Age IX ; nor do we find that all leptomorphic individuals are schizothymic or that pyknic individuals are nearly always cyclothymic. A correlation technique therefore becomes essential.

With qualitative variables as with quantitative, the most useful factors will be those that lead to a hierarchical classification ; and the natural procedure will be to seek, in any particular ' domain ' or ' universe of discourse,' first the factor which will account for the largest amount of individual variation, and then the factor which will account for the largest amount of residual variation, and so on. Thus, to adopt an expressive term familiar to those brought up on Aristotelian logic, a factor may be thought as a διαφορὰ εἰδοποιός [2]—a hypothetical ' differentiating principle,

[1] E.g., the contributions by different writers in Borel, E. (ed.) *Traité du Calcul des Probabilités* (1924 onwards). Kelley remarks on the fact that, whereas earlier writers who touched upon the statistical treatment of psychological problems classified ' statistical series ' into ' qualitative ' and ' quantitative,' more recent writers (like Garrett, Thurstone, Engelhart, McNemar) have preferred to classify series into ' discrete ' and ' continuous,' and that, as they use the terms, the distinctions are by no means identical. Quantitative scales may also be discrete ; and, so far as such writers touch upon discrete distributions, they restrict themselves almost entirely to quantitative instances (cf. 13, pp. 62f.).

To avoid misunderstanding, let me emphasize at once that I do not for one moment believe that all the psychologist's observations can be adequately dealt with by just tabulating frequencies. In my view *the method of trait-assessment, whether qualitative or quantitative, should always be supplemented by the method of case-study.* Neither is complete without the other. The ancient controversy between ' clinical ' and ' metrical ' psychologists has arisen largely because each seems to have assumed that his own method was the sole method. How the two kinds of approach may be combined in a single research I endeavoured to show in my book on *The Young Delinquent.* And I should hold that almost every problem of individual psychology demands this double mode of treatment.

[2] Arist., *Top.* VI, vi, 2 ; cf. *Nicom. Eth.*, X, iv, 3.

constituting a species.' Remembering that 'species' is a relative term, we shall therefore expect the widest factor to represent the genus under investigation; and the remaining factors to represent the differences between species, sub-species, and so on, in hierarchical order. However, as modern logicians have so often pointed out, the traditional notion of classification found in the older text-books of logic assumed a neatly arranged system of natural kinds—each genus sharply distinguished from all other genera, each species or type unequivocally separated from the rest; whereas in actual fact there is a vast and often irregular amount of overlapping, and this overlapping is greatest in the more complex sciences. Hence, in my view, the only safe and speedy way to reach a satisfactory solution in such cases is to invoke the aid of a suitable statistical technique. And, if the statistician has not as yet developed an appropriate method which he can offer us, then the psychologist, however imperfectly, must set to work to devise his own.[1]

II. ALTERNATIVE STATISTICAL TECHNIQUES

Correlation in the Case of Two Dichotomous Classifications. In my first attempt at factorizing results obtained with the Binet scale, each test was paired with every other, and the number of children passing both was taken as indicating the association between the abilities so tested. In view of the large number of tests in the scale, the frequencies for the several pairs were entered in a single table of double entry, like an ordinary contingency-table: the whole table was then fitted with theoretical values by means of the simple summation formula; and the sums were taken as indicating the degree to which each test was representative of the common factor.[2] In later researches, when the factor analysis of correlation coefficients had become a recognized procedure, attempts were made to reduce the initial data to correlational form. Since cognitive abilities are continuous variables, the ideal procedure, so it was thought, was to modify the Binet problems, wherever possible, in order to secure graded measurements, and, where this was not possible, to calculate

[1] The logical difficulties surrounding attempts at classification in the more complex sciences are perhaps most clearly brought out by Sigwart's discussion; and his solution is in keeping with the procedure suggested above. By what principle, he asks, are we to be guided in our "progressive efforts to work by ascending generalization from the lower to the higher classes, or downwards from the higher to the lower"? At each stage, he replies, "the most serviceable arrangement will be that which groups together those things which have the largest number of characteristics in common" (*Logic*, Part III, 'Technical'; esp. pp. 158f.: it is suggestive to note that he proceeds to cite, as the cases showing the most difficult and complicated structure, the attributes or 'elementary characteristics' studied in psychology).

But how (he goes on to ask) are we to compile our preliminary list of characteristics in order to see what is the largest number that certain groups have in common? That, as he points out, is plainly not a formal but a material problem. To which I need only add that, if it is a material problem, then it is not a problem that can be decided by statistical calculation alone.

There is, I am well aware, a widespread notion that, in dealing with quantitative measurements, the investigator can proceed by first taking all the tests that are applicable or throwing together all data that may be available, trusting that the statistical analysis will then "discover" the underlying factors. But that, I venture to assert, is quite fallacious. The initial collection of data is or should be guided by some prior hypothesis which the factor analysis is intended to check; and the very process of collecting the data must imply some tentative ideas, conscious or unconscious, as to the nature of the classification which the investigator is testing. Most probably these ideas have resulted from some causal theory that has led him to suspect that these or those distinctive manifestations may be grouped together to form this or that species or genus. For example, in psychology neurological knowledge may have led him to anticipate a general cognitive factor; or anthropological knowledge to expect a conjunction of certain shades of hair, eye, and complexion; or again his psychiatric theories or his case-studies may have induced him to believe that certain signs or symptoms will cluster together. But whatever they may be, all such prior considerations will, as a rule, be based not on a statistical analysis, but on concrete observations or on speculative hypotheses.

[2] Cf. Yule, U., *loc. cit.*, p. 64, eq. 1, and p. 66, Table III. It was in fact this formula which suggested the use of 'simple summation' and the application of the chi-squared test of significance for correlation tables calculated from quantitative variables in the ordinary way.

tetrachoric coefficients, on the ground that the distribution of such abilities conforms pretty closely with the normal curve.[1]

However, as I have already pointed out, with much of the data obtained in work on non-cognitive differences, the information may be essentially qualitative in nature and available only in classificatory shape. In such cases the tetrachoric coefficient is quite unsuitable. Nevertheless, without making any specific assumptions about the nature or distribution of the underlying variables, it is still possible to express the association between such characteristics by means of correlation. Thus, to obtain what in form at least is a product-moment coefficient, Yule [2] suggested taking the line of division between each pair of classes as the arbitrary origin and the class interval as unit, and then calculating the mean, the standard deviations, and the product-sum in the usual way. Using the notation I have adopted in previous articles,[3] suppose that, in a population of $N = N_0 + N_1$, a given attribute is present in N_1 cases and absent in N_0 cases; then the mean will be $\frac{1}{2N}(N_1 - N_0) = \frac{N_0}{N} - \frac{1}{2}$, and the standard deviation $\frac{\sqrt{N_0 N_1}}{N}$. This is tantamount to assigning a value of $+\frac{N_0}{N}$ (the mean of the positive subgroup) to denote the *presence* of the attribute, and $-\frac{N_1}{N}$ (the mean of the other subgroup) to denote its *absence*. We can now compute the frequencies for joint deviations; and, with this or some equivalent device, we can always calculate a complete set of correlation coefficients between all possible pairs from a series of attributes. Then, if we wish, the table of coefficients can be factorized, and the attributes expressed in terms of the factors so obtained. This method—namely, the calculation of coefficients from all the 2×2 tables formed from the initial tables of frequencies—was adopted for a factorial study of the Binet tests, and appeared to give quite satisfactory results.[4]

Available Procedures. Evidently, then, there is no intrinsic difficulty about factorizing data of this kind. According to the problem at issue or the assumptions we prefer to make, various working methods may be adopted. From the standpoint of the research-student, perhaps the simplest way to classify the possible procedures is to begin with the familiar distinction between (A) methods appropriate to cases where *measurement* is at least conceivable and (B) methods appropriate to cases where it is more natural to think in terms of *frequencies*.

(A) Methods of the first type proceed by converting the initial data, explicitly or implicitly, into quantitative form, and then base the factor analysis on a table of correlations or of some equivalent type of coefficient. For the present purpose the most useful coefficients are the following:

(i) the product-moment correlation for a twofold point distribution (φ);
(ii) Yule's colligation coefficient (φ for an equalized table);
(iii) Yule's association coefficient;
(iv) Pearson's coefficient of contingency, corrected if necessary for degrees of freedom by Tschuprow's formula;
(v) Pearson's tetrachoric coefficient;
(vi) the correlation ratio.

[1] Cf. John, Enid, and Burt, C., 'Factor Analysis of the Terman Binet Scale,' *Brit. J. Educ. Psych.*, XII, 1942, p. 122. On the relative merits of different methods of calculating association coefficients for such data as are furnished by the Binet tests, cf. *Mental and Scholastic Tests*, pp. 137 and 232, also *The Backward Child*, Appendix III. As to the reasons for preferring graded measurements, see *id.*, 'The Measurement of Intelligence by the Binet Tests,' *Eugenics Review*, IV, 1914, pp. 140f.

[2] Yule, U., *Introduction to the Theory of Statistics*, 1912, pp. 216–7.

[3] E.g., Burt, C., 'Statistical Problems in the Evaluation of Army Tests,' *Psychometrika*, IX, 1944, p. 225, and refs.

[4] Cf. *Mental and Scholastic Tests*, p. 137. As I noted at the time, however, the use of an association coefficient or point-distribution correlation commonly introduces "an additional factor which seems to indicate the relative difficulty of the various tests' (*loc. cit.*, footnote 1). This factor, as a rule, does not appear when tetrachoric or ordinary product-moment coefficients are used, because they eliminate differences in the point of division.

(B) Methods of the second type leave the data in qualitative form. In dealing with such material we should, I think, frankly recognize that factor analysis may be legitimately applied to any kind of symmetrical table, not merely to correlations or covariances. In this case the figures to be factorized will mostly take one or other of the following forms :

(i) the absolute frequencies ;
(ii) the relative frequencies or probabilities ;
(iii) some other ratio, e.g., the square contingency or its derivatives.

To a large extent the distinctions on which the above classification rests are distinctions of procedure rather than of principle. Both the contingency coefficient and the root-mean-square contingency (φ) are closely related to the square contingency (χ^2) and to the frequencies on which it is based ; and thus the two main types of procedure tend to merge the one into the other.

III. THE TREATMENT OF MULTIPLE DETERMINATES

The General Table of Observed Data. With the Binet-Simon scale, and, indeed, with most of the qualitative tests which the psychologist is likely to factorize, the presence or absence of the ' attribute ' yields in each case a dichotomous or ' bipolar ' classification of the individuals observed. For the majority of such cases I have suggested that the φ-coefficient can be regarded as the simplest and most convenient procedure.[1] However, in constructing contingency-tables, it is often essential to consider cases in which the classification is (to preserve Yule's terminology) not ' twofold,' but ' manifold.' [2] Tables of this type, and particularly the compound table, where there are not only more than two possibilities for each trait, but also more than two traits (a type of table which Yule barely mentions), have been much neglected in the past ; and it is one of the chief objects of this paper to direct attention to the problems which it raises. Evidently a twofold classification may always be regarded as a special and simple case of a manifold classification. Hence it will be more instructive to discuss the question in its most general form.

In a ' manifold ' classification " the individuals are first divided under n_i heads, A_1, A_2, . . . A_{n_i} " ; this classification may then be compared with a number of other classifications, dividing the same individuals into the classes B_1, B_2, . . . B_{n_j}, and so on (6, p. 60). To meet the more complex cases I propose to introduce a terminology familiar to logicians, and call the general character to which the classification relates (*i.e.*, what scholastic writers termed the *fundamentum divisionis*) a ' determinable,' and the alternative qualities assigned to the determinable the ' determinates.' [3] Thus, to take Yule's example, hair-colour may be called a ' determinable,' and the various shades that hair-colour is recorded as taking—' fair,' ' red,' ' dark '—the ' determinates.' Accordingly, the matrix of observed data which we

[1] The problem arose afresh in connexion with the analysis of data accumulated by psychologists working in the Directorate of Personnel Selection at the War Office during the war. Much of the present section is based on a Memorandum which I was asked to submit in 1943. For the special case in which the data yield simple dichotomous classifications I there suggested the procedure described in *Mental and Scholastic Tests*, namely, substituting a point-distribution calculated from the resulting 2×2 tables in place of the absolute or relative frequencies. A slightly different mode of approach was suggested by Mr. P. Slater, which he has since described more fully in a suggestive paper on ' The Factor Analysis of a Matrix of 2×2 Tables ' (*J. Roy. Stat. Soc.*, IX (Sup.), 1947, pp. 114–27). Apart from theoretical objections, I am inclined to think that the working procedure would in practice be found somewhat laborious. Moreover, most of the traits with which he is concerned (" shyness," " apprehensiveness," " emotional instability " and the like) seem to me to imply *graded* characteristics, which (so far as is known) have an approximately normal distribution. Hence for these the tetrachoric coefficient would seem to be more appropriate.

[2] *Loc. cit.*, p. 60.

[3] I have borrowed the terms from W. E. Johnson, *Logic* (1921), Pt. I, Chap. XI, p. 174.

require to investigate will have the form [1] shown in Table I. Let us call this matrix M_1, since it corresponds to the non-standardized matrix of observed measurements in quantitative investigations.

TABLE I. OBSERVED TABLE SHOWING PRESENCE OR ABSENCE OF ATTRIBUTES FOR A GIVEN SAMPLE OF PERSONS

	Persons (N)					Total for Row
	Tom	Dick	Harry	...	George	
Determinables (m)						
Hair-colour						
Fair	1	0	0	...	0	N_1
Red	0	1	0	...	0	N_2
Dark	0	0	1	...	1	N_3
Subtotal	1	1	1	...	1	N
Eye-colour						
Light	1	0	1	...	0	N_4
Brown	0	1	0	...	1	N_5
Subtotal	1	1	1	...	1	N
Total for Column ..	m	m	m	...	m	Nm

It will be convenient to adopt the following notation : [2]

Number of determinables	m,
Number of determinate values in the 1st, 2nd, ... jth, ... mth determinable	$n_1, n_2 \ldots n_j, \ldots, n_m$,
Total number of determinates	$n = n_1 + n_2 + \ldots + n_m$
Total number of persons for each determinate ..	$N_1, N_2, \ldots N_{nj}$,
Total number of persons in sample	$N = N_1 + N_2 + \ldots + N_{n_j}$,
Number of persons possessing both the jth and the kth determinates	N_{jk}.

Since for each determinable every person must possess one determinate value and none can possess more than one (e.g., every person must have hair of *some* colour—either fair, dark, or red), the subtotal obtained by each person for each determinable must be 1. It follows that, if for the first determinable (say) the number of rows is n_1, then, when we know the values for $(n_1 - 1)$ of those rows and knowing the values for the row of totals (I), we can immediately deduce the value for the remaining row. Accordingly, assuming $N > n$, the rank of the matrix represented by Table I will be $(n_1 - 1) + (n_2 - 1) + \ldots + (n_m - 1) + 1 = n - m + 1$.

[1] Since for each determinable the last row of frequencies can be deduced from the preceding rows, we may, if we prefer, omit it from the subsequent calculations : that would evidently give rise to a slightly different form of analysis, but the modifications entailed will be obvious. When every one of the determinables permits of two alternatives only (as in dealing with characteristics recorded as ' present ' or ' absent,' or with test-items that are marked only ' pass ' or ' fail '), and when we record only one of the two determinates, the table reduces to the form originally used for the Binet tests. Let me add that, in certain cases, it is desirable to work with compound characteristics. Thus not infrequently the mere presence of one or other of two characteristics may have no diagnostic significance by themselves, but their joint presence may be highly significant. For example, in an adolescent girl, pallor by itself has little importance, and increasing plumpness still less ; but the two together will at once be suggestive of chlorosis. Similarly in emotional and moral disturbances, it is the total syndrome that is of diagnostic importance, not the isolated symptom. For statistical purposes, any such compound characteristic should still be treated as a single attribute, even though it involves a cluster of traits.

[2] Here and later, instead of Yule's notation for contingency-tables, I shall keep, so far as possib e, to that introduced by Kelley and adopted by most statistical psychologists in this country cf. Kelley, T. L., *Statistical Method*, 1923, pp. 263f.).

The Unstandardized Product-sum Matrix. By post-multiplying the table of observed frequencies by its transpose, we obtain a product-sum matrix

$$C_t = M_1 M_1' . \qquad \text{(i)}$$

Except for the fact that the values in M_1 are not standardized deviates, this product-sum matrix corresponds with the more familiar matrix of covariances (or rather codeviances) between *traits*. We can obtain an alternative product-sum matrix by pre-multiplying M_1 by its transpose, namely, $C_p = M_1' M_1$. The result will correspond to the matrix of covariances or codeviances between *persons*. This second type of matrix, however, will not concern us at this stage.

The matrix of product-sums for traits constitutes a symmetrical contingency-table. Consider, for example, the calculation of the product-sum for a pair of determinates from two different determinables, such as those set out below :

															Total
Hair, fair (j)	1	1	1	1	0	1	0	1	1	0	1	0	1	1	0	N_j
Eyes, light (k)	1	1	1	0	1	1	1	0	1	0	0	1	1	1	1	N_k
Product..	1	1	1	0	0	1	0	0	1	0	0	0	1	1	0	N_{jk}

Evidently the product-sum, $N_{jk} = 7$, states the number of persons who possess *both* traits (a ' second order ' frequency) ; and this is the value that would be entered in one cell of a small contingency-table that might be drawn up to show the relation between hair-colour and eye-colour. (For an illustration of the type of table eventually obtained, see p. 177.)

However, as the size of the sample (N) will necessarily differ from one investigation to another, it will be better to express the data in terms not of the absolute but of the relative frequencies, e.g., $p_{jk} = N_{jk}/N$ say. We can then go on to standardize the data by dividing this figure by the standard deviation, $\sqrt{N_j N_k / N}$. In this way we get a set of proportionate values which we may designate

$$r_{jk} = p_{jk} / \sqrt{p_j p_k} = N_{jk} / \sqrt{N_j N_k} . \qquad \text{(ii)}$$

This is equivalent to pre- and post-multiplying the product-sum matrix by a diagonal matrix $D^{-\frac{1}{2}}$, whose elements are $1/\sqrt{N_j}$. We thus arrive at a modified symmetrical matrix

$$R_t = D^{-\frac{1}{2}} C_t D^{-\frac{1}{2}} = D^{-\frac{1}{2}} M_1 M_1' D^{-\frac{1}{2}} = MM' \text{ say} . \qquad \text{(iii)}$$

As will be shown in a moment, this matrix will be positive semidefinite ; being symmetrical, it will therefore have Gramian properties.[1] And further we may note that each element in its leading diagonal now has the value of unity. The matrix may thus be regarded as in standard form. (See Table III, p. 178.)

Factor Analysis. We may now proceed to factorize this standardized symmetrical matrix by finding its latent roots and latent vectors. This will enable us to express our $n \times n$ observed matrix in terms of a set of r orthogonal factors (where $r = n - m + 1$). Adopting the usual notation we shall have

$$D^{-\frac{1}{2}} M_1 = M = L V^{\frac{1}{2}} P = FP , \qquad \text{(iv)}$$

where P will be an $r \times N$ section of an orthogonal matrix ($PP' = I$), L an $n \times r$ section of another orthogonal matrix ($L'L = I$), and $V^{\frac{1}{2}}$ an $r \times r$ diagonal matrix. This gives

$$R_t = MM' = LV^{\frac{1}{2}} P . P' V^{\frac{1}{2}} L' = LVL' , \qquad \text{(v)}$$

or (on taking product-moments for persons rather than traits)

$$R_p = M'M = P' V^{\frac{1}{2}} L' . LV^{\frac{1}{2}} P = P'VP . \qquad \text{(vi)}$$

[1] If all the principal minors of a matrix are non-negative, i.e., greater than or equal to zero, it is said to be positive definite (or semidefinite, if its rank is less than its order). If in addition it is symmetrical, it is said to be Gramian. A related but less general property, which characterizes many contingency-tables of the kind I have described, is that which Yule terms ' isotropy ' ; its presence has an obvious interest for the factorist : (see 6, pp. 68–70, and refs., p. 73).

In either case V will be the diagonal matrix containing latent roots of both R_t and R_p, and L and P will be the latent vectors of R_t and R_p respectively. To obtain appropriate equations for calculating P from M, say, $W'M = P$, we have $L'M = L'LV^{\frac{1}{2}}P = V^{\frac{1}{2}}P$.

Hence $$P = V^{-\frac{1}{2}}L'M\,; \qquad \text{(vii)}$$

This suggests that we may put $W' = V^{-\frac{1}{2}}L' = V^{-1}F$.

Values for V and F can be computed by following the usual procedure. By hypothesis the first row of P will be a set of hypothetical measurements, one for each person, and will be so calculated that, when multiplied by the first column of the weighting matrix F, the product will be an $n \times m$ hierarchical matrix giving the closest possible fit to M. Let f be the first column of the weighting matrix. The familiar method for solving the equations to obtain f and v leads to the introduction of v as a Lagrange multiplier. The conditions for minimizing the squared discrepancies then yield an equation of the well-known form

$$f'(R_t - vI) = 0\,, \qquad \text{(viii)}$$

which will only be satisfied if

$$|R_t - vI| = 0\,. \qquad \text{(ix)}$$

The solution yields r values for v. To obtain the first factor, we take the largest value for v, insert it in equation (viii), and solve for f'. Then, taking the other values for v in order of magnitude, we find the remaining rows for F'. In practice the values can be calculated by the iterative method known as ' weighted summation.'

Interpretation of Factorial Results. It will be seen that the calculations I have described closely resemble those proposed by Pearson for analysing a correlated set of *quantitative* traits, such as height, arm-length, leg-length, etc. But, since our starting point has been a table of frequencies rather than a table of measurements, we cannot press the analogy too far without first considering in greater detail what precisely is the nature of the results thus reached.

Equation (iv) may be interpreted by saying that the modified matrix of observed data ($M = D^{-\frac{1}{2}}M_1$) has been expressed in terms of two factor-matrices : (*a*) a matrix of ' uncorrelated ' factor-measurements for *persons*, P, so chosen that the first factor will account for the maximum amount of the variance exhibited by the initial matrix, and each successive factor will account for a maximum amount of the variance exhibited by each successive residual matrix ; and (*b*) a matrix of ' uncorrelated ' factor-saturations for the several *traits*, giving the weights to be applied to each factor in reconstructing the observed trait-measurements. If we prefer to consider ourselves as factorizing the unmodified matrix of observed data, F, we can write equation (iv) in the form $M_1 = (D^{\frac{1}{2}}F)P$, and $D^{\frac{1}{2}}F = D^{\frac{1}{2}}LV^{\frac{1}{2}}$ will now represent the factor-saturations or weights ; but the columns of factor-saturations ($D^{\frac{1}{2}}F$) will no longer be uncorrelated.

Moreover, if in equation (viii) we substitute for R_t from equation (iii), we obtain

$$f'(D^{-\frac{1}{2}}M_1M_1'D^{-\frac{1}{2}} - vI) = 0\,. \qquad \text{(x)}$$

We can simplify this by writing $f'D^{-\frac{1}{2}} = u'$ and post-multiplying by $D^{\frac{1}{2}}$. We thus obtain

$$u'(M_1M_1' - vD) = 0, \text{ that is, } u'M_1M_1' = vu'D\,. \qquad \text{(xi)}$$

We now see that one solution of this equation will be obtained by taking $u' = [1, 1, \ldots 1] = w_0'$ (in the usual notation), and $v = m$. For $w_0'M_1 = [m, m, \ldots m]$ (i.e., the sums of the several columns in Table I) $= m\,w_0'$; and $w_0'M_1' = [N_1, N_2 \ldots N_n]$ (i.e., the sums of the several rows of the same table) $= w_0'D$. We thus obtain

$$u'M_1M_1' = m\,w_0'M_1' = m\,w_0'D = vu'D\,,$$

in accordance with equation (xi). This procedure will yield only the proportionate values for the latent vectors, in this case $f' = w_0'D^{\frac{1}{2}} = [\sqrt{N_1}, \sqrt{N_2}, \ldots \sqrt{N_n}]$. But since the elements of l' are also proportional to those of f', and $l'l = 1$, we can first calculate values of l' by normalizing the above values for f'. This gives

$$l' = \left[\sqrt{\frac{N_1}{N_m}}, \sqrt{\frac{N_2}{N_m}}, \ldots \sqrt{\frac{N_n}{N_m}}\right], \text{ since } N_1 + N_2 + \ldots + N_n = N_m\,.$$

Now $f' = v^{\frac{1}{2}}l' = \sqrt{m}.l'$. Hence we can at length obtain appropriate values for the saturations namely,

$$f' = \left[\sqrt{\frac{N_1}{N}}, \sqrt{\frac{N_2}{N}}, \ldots \sqrt{\frac{N_n}{N}}\right]\,. \qquad \text{(xii)}$$

It is easy to show algebraically that the expression thus reached for the saturations conforms with the requirements for a first ' principal axis.' The requirements can be formulated quite simply.

Since $f'(R - v\,I) = 0$, we must have $f'\,R = v\,f'$: i.e., if f' denotes a row of saturations for a principal axis of the symmetrical matrix R, the row of weighted column sums obtained by post-multiplying the saturations by the matrix R must be proportional to the original row of saturations, and the constant of proportionality must be $f'f = v$, the factor-variance or 'latent root.' Accordingly, consider the section of the ith column of R for the first determinable. This has the following form :

$$\frac{N_{1j}}{\sqrt{N_1 N_j}}$$
$$\frac{N_{2j}}{\sqrt{N_2 N_j}}$$
$$\frac{N_{3j}}{\sqrt{N_3 N_j}}$$
$$\cdots$$

On multiplying each row of R by the saturation coefficient, we obtain

$$\frac{N_{1j}}{\sqrt{N_1 N_j}} \cdot \sqrt{\frac{N_1}{N}} = \frac{N_{1j}}{\sqrt{N_j N}}$$
$$\frac{N_{2j}}{\sqrt{N_2 N_j}} \cdot \sqrt{\frac{N_2}{N}} = \frac{N_{2j}}{\sqrt{N_j N}}$$
$$\frac{N_{3j}}{\sqrt{N_3 N_j}} \cdot \sqrt{\frac{N_3}{N}} = \frac{N_{3j}}{\sqrt{N_j N}}$$
$$\cdots \qquad\qquad \cdots$$

But the total number showing these determinate values is N_j (i.e., if there are only three such values, $N_{1j} + N_{2j} + N_{3j} = N_j$). Hence the total of the column will be $\frac{N_j}{\sqrt{N_j N}} = \sqrt{\frac{N_j}{N}}$. If there are m determinates, the total for each column, after every row has been weighted by the appropriate saturation coefficient, will be $m\sqrt{\frac{N_j}{N}}$. Dividing by m (the latent root), we obtain the saturation coefficient. The required condition is thus fulfilled.

Since the number of factors in R (i.e., the rank of R) is less than the number of determinate traits (i.e., the order of R), each of the remaining latent roots will have a value that is less than m. Hence the factor just extracted is the factor which accounts for the maximum amount of the total variance, and may therefore be regarded as the first or general factor.

From the values given for the saturations by equation xii, we can now reconstruct the symmetrical hierarchy due to the first factor, namely, $H = ff'$. We obtain the symmetrical matrix

$$\begin{bmatrix} \frac{N_1}{N} & \frac{\sqrt{N_1 N_2}}{N} & \cdots & \frac{\sqrt{N_1 N_n}}{N} \\ \frac{\sqrt{N_2 N_1}}{N} & \frac{N_2}{N} & \cdots & \frac{\sqrt{N_2 N_n}}{N} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

We may write the general element in this matrix $h_{ij} = \frac{\sqrt{N_i N_j}}{N}$. But the general element in the standardized product-sum matrix was $r_{ij} = \frac{N_{ij}}{\sqrt{N_i N_j}}$. Hence the general element in the residual matrix will be

$$r_{ij} - h_{ij} = \frac{N_{ij}}{\sqrt{N_i N_j}} - \frac{\sqrt{N_i N_j}}{N}$$
$$= \frac{1}{\sqrt{N_i N_j}} \left(N_{ij} - \frac{N_i N_j}{N} \right).$$

Now $\frac{N_i N_j}{N}$ is the 'expected value' of the frequency in the ijth subgroup. This 'expected value' is the frequency we should deduce for that subgroup if it consisted simply of individuals.

sampled at random from the general population represented by the N persons whom we have observed, and if there were no special correlation between the ith and the jth traits, due to some causal influence affecting those traits. Using a notation that has recently come into popular use [1] let us write $\frac{N_i N_j}{N} = E$ for the expected frequency, and $N_{ij} = O$ for the frequency actually observed. Then $N(r_{ij} - h_{ij})^2 = \frac{(E - O)^2}{E}$; and thus expresses the square of the discrepancy between the observed value and the expected value for any element in the product-sum matrix. Now, if such discrepancies were due solely to chance, we should expect the sum of their squares to be distributed in accordance with chi-squared. In fact, $\chi^2 = \Sigma \frac{(E - O)^2}{E}$ is the value we should calculate if we wished to test the set of observed data to determine whether or not they represented the results of random sampling from the general population.[2]

The meaning of what we have called the saturations for the first factor is obvious. In factor analysis the first or ' general ' factor describes the general quality defining the entire *genus* with which we are dealing. Its saturations express the relative importance of the several traits as manifestations of that generic factor. Here they are actually proportional to $\sqrt{N_1}$, $\sqrt{N_2}$, . . . $\sqrt{N_n}$, i.e., to the square root of the frequency with which the several traits occur in the population sampled. The factor-measurements for each individual can be calculated (if required) in accordance with the usual procedure. They will evidently show how far that person is typical of the general population represented by the sample.

The factor-saturations for the remaining factors can be extracted in the usual way and tested by the chi-squared procedure.

The Conditions of Consistency. Figures descriptive of one and the same population or sample must be mutually compatible, and therefore obey certain restrictive conditions. The precise formulation of such conditions has a double value : it enables us (i) to detect data that are inaccurate and (ii) to assign limits to the unknown values when the data are incomplete. Elsewhere I have ventured to suggest that, whether the traits concerned are qualitative attributes or quantitative variables,[3] the most general criterion is essentially the same, namely, that *each of the principal minors of the symmetric table of product-sums shall be non-negative.*

In *Laboratory Notes on Factor Analysis* this principle was deduced for quantitative variables ; it remains to show that it also holds good for qualitative data. The arguments are much the same in either case. Let R_i denote a symmetric submatrix, obtained from R (eq. iii) by striking out corresponding rows and columns for certain variables. Then $|R_i|$ will be a principal minor of R. Now R_i, like R, can be factorized into its latent vectors and roots. Thus, we can write $R_i = L_i V_i L_i'$, and therefore $|R_i| = |L_i||V_i||L_i'| = |V_i|$ (since L_i is orthogonal and $L_i = 1$). When the rank of R_i (r_i, say) is equal to its order (n_i), then $V_i = F_i' F_i$ and forms a diagonal matrix in which each element consists of the sum of the *squares* of the saturations. When $r_i < n_i$, the extra elements in V_i will be zero. Hence, whatever the rank of R, no element in V_i can be negative ; therefore the value of $|R_i|$ is also non-negative. This clearly holds for *any* principal minor in R, and consequently for any that can be derived from C.

In practice it is unnecessary to test *every* principal minor. By invoking the Binet-Cauchy theorem, it can readily be proved that, if the *leading* principal minors of the 1st, 2nd, . . . rth order are positive, then *all* the principal minors of those orders will be positive. Now the former are the minors which automatically appear, stage by stage, when we adopt what I have called the ' factorial method ' of calculating partial correlations by hierarchical subtraction.[4] And it is easy to show that,

[1] E.g., by McNemar (*Psychological Statistics*, 1949, pp. 179, 186), and other recent text-books.

[2] It may be noted that here as in an analysis of variance the averages (means) are not eliminated from the very outset as though they were devoid of interest, and as they would be in an ordinary correlational procedure. They are, in effect, tested for significance.

[3] In the case of *qualitative attributes*, the working criterion suggested by Yule is that no ultimate class frequency can be negative ; in the case of *quantitative variables* his basic principle is that no partial correlation can exceed the numerical value of unity : cf. (6), Chap. II, pp. 17–24, 250–2, and refs. ; also *id.*, ' On the Theory of Consistence of Logical Class Frequencies,' *Phil. Trans. A*, CXCVII, 1901, pp. 91f.

[4] See *Laboratory Notes*, sect. 8, ' The Solution of Simultaneous Equations ' : (a brief summary of the method, with a worked example, will be found in Vernon's *Notes on Statistical Method*, pp. 42f.).

if the condition specified is fulfilled, then (i) in the case of qualitative data all frequencies up to those of the order included in the product-sum matrix must have real and positive values, (ii) in the case of quantitative data no partial correlation can exceed the numerical value of 1, and (iii) in either case the tables can be analysed in terms of real factors. It will be sufficient to indicate quite briefly the general lines of proof for these three statements.

(i) *Higher Frequencies.* Since the square-sums of frequencies are necessarily larger than their product-sums, the appearance of a negative principal minor would imply that some of the frequencies had negative or imaginary values. If the product-sum matrix includes only frequencies of the second order, the condition as stated is necessary but not sufficient. To show that the *ultimate* frequencies have real and positive values, it would be necessary to introduce composite attributes.

(ii) *Partial Correlations.* No generality will be lost if we suppose the correlation matrix or product-sum to be arranged so that the rows and columns for the two correlated variables shall be the last. Any variables which are neither to be correlated nor to be partialed out will be deleted from the matrix. Then, the partial correlation,

$$r_{(n-1)n.12\ldots(n-2)} = r \text{ (say)} = \frac{D_{(n-1)n}}{\sqrt{D_{(n-1)(n-1)}}\sqrt{D_{nn}}}, \quad . \quad . \quad . \quad \text{(xiii)}$$

where D, the determinant obtained from R after the irrelevant rows and columns have been deleted, will be a principal minor of R. Now, if $1 \geqslant r^2$, evidently $D_{(n-1)(n-1)}\, D_{nn}$ must be $\geqslant D^2_{(n-1)n}$; that is

$$\begin{vmatrix} D_{(n-1)(n-1)}, & D_{(n-1)n} \\ D_{(n-1)n} & , D_{nn} \end{vmatrix} = \text{non-negative.} \quad . \quad . \quad . \quad \text{(xiv)}$$

Now

$$D = \begin{vmatrix} & \\ & D_{n(n-1),\,n(n-1)} \\ & \end{vmatrix}^{-1} \begin{vmatrix} D_{(n-1)(n-1)}, & D_{(n-1)n} \\ D_{(n-1)n} & , D_{nn} \end{vmatrix}, \quad . \quad . \quad . \quad \text{(xv)}$$

where the divisor must be positive. Hence, if D (the principal minor) is non-negative, then the expression on the left of (xiv) must be non-negative, and we have $r^2 \leqslant 1$; if D is negative, then $r^2 > 1$.

(iii) *Real Factors.* Now, whatever factorial methods we adopt to reduce R to a diagonal matrix, V say, we can split this diagonal matrix, and so further reduce R to the form $R = F\,I\,F'$, where $F = LV^{\frac{1}{2}}$. But this further process will in general be non-rational [1]; and if any of the non-zero elements in V (and therefore in I) are negative, $V^{\frac{1}{2}}$ and consequently F will involve non-real elements. But, by Sylvester's law of inertia, however we reduce R to the form $F\,I\,F'$, the number of negative values in I must remain the same. Therefore, if either of the methods of factorization described above yields imaginary elements for the factor matrix F, then *any* method must yield such elements: that is, it will be impossible to analyse R into factors with real elements. Conversely, if our criterion is fulfilled, then R will be reducible to terms of real factors.

We may therefore sum up our conclusions regarding the table of product-sums as follows: in order that an $n \times n$ table shall be (*a*) self-consistent and (*b*) analysable in terms of r *real* factors ($r \leqslant n$), it is at once necessary and sufficient that *all its leading principal minors shall be non-negative.*

IV. A FACTORIAL ANALYSIS OF PHYSICAL ATTRIBUTES

The Neglect of Qualitative Traits in Studies of Body Types. To illustrate the general procedure proposed may I now describe its application in a simple concrete case? The various points in my argument will perhaps be most readily grasped if I choose a problem from physical anthropology. The characters in question will then be easier to visualize; less doubt is likely to arise about which characters are genuinely qualitative; and at the same time it will be possible to show how quantitative measurements may be included in the same series and subjected to the same general treatment.

Recent discussions about the relation between temperamental and physical characteristics have been limited almost entirely to those physical characteristics which, like height and weight, may be treated as continuously graded, quantitative variables; they have tended almost entirely to overlook the possible relevance of qualitative characteristics, such as eye-colour, hair-colour, or peculiarities in the shape of certain features, like the nose or chin (cf. 10 and refs.). In a later paper I hope to examine what verifiable support there may be for popular and traditional notions about the value of such traits as a clue to individual character. Here my purpose is primarily to illustrate a practicable technique for expressing, in terms of 'factors' or type-tendencies, any significant associations that may be observable among the physical traits themselves.

[1] Cullis, C. E., *Matrices and Determinoids*, II, p. 293, Theorem III, and *Factors of the Mind*, p. 307, footnote 1.

The Data. In an early investigation which I commenced some years ago, I endeavoured to determine more particularly whether there was any evidence for such associations in the case of those traits that were commonly supposed to characterize certain 'racial' types (cf. 7, pp. 23, 27, and refs.). For this purpose I collected observations on the characteristics most frequently used in classifying individuals into such groups, namely, eye-colour, hair-colour, stature, shape of head, shape of nose and chin, and type of hair. Here it will be unnecessary to defend or define the various terms adopted: as regards stature it will be sufficient to say that persons were classed as 'tall' if their height was above the average for the general population [1] and 'short' if it was below the average. Under 'wide' heads I here include all with a cranial index over 75 (that is, mesocephalic heads, as well as dolichocephalic in the stricter sense). To shorten the table, I have also pooled certain traits and omitted others.

The data were collected at Liverpool, a district where representatives of different nationalities—Welsh, Irish, Scots, as well as foreigners—were easily accessible. In most cases temperamental assessments were obtained at the same time; but these will also be omitted from the present tables, as they raise somewhat different issues. In all, 217 individuals were examined, about two-thirds of them males. But, partly to simplify the calculations and partly because the later observations were rather more trustworthy, I shall here restrict my analysis to the data obtained from the last hundred males in the series.

The Crude and Standardized Contingency-Tables. The number of persons characterized by the several attributes specified, taken in pairs, are set out in Table II (this may be taken as representing the initial type of contingency-table which was designated C_t on p. 172 above).

TABLE II. OBSERVED FREQUENCIES
Number of Persons exhibiting Characteristics Specified

Trait	F	R	D	Tot.	L	M	B	Tot.	N	W	Tot.	T	S	Tot.
HAIR														
Fair ..	22	0	0	22	14	6	2	22	14	8	22	13	9	22
Red ..	0	15	0	15	8	5	2	15	11	4	15	10	5	15
Dark ..	0	0	63	63	11	25	27	63	44	19	63	20	43	63
Total ..	22	15	63	100	33	36	31	100	69	31	100	43	57	100
EYES														
Light ..	14	8	11	33	33	0	0	33	27	6	33	29	4	33
Mixed ..	6	5	25	36	0	36	0	36	20	16	36	10	26	36
Brown ..	2	2	27	31	0	0	31	31	22	9	31	4	27	31
Total ..	22	15	63	100	33	36	31	100	69	31	100	43	57	100
HEAD														
Narrow..	14	11	44	69	27	20	22	69	69	0	69	30	39	69
Wide ..	8	4	19	31	6	16	9	31	0	31	31	13	18	31
Total ..	22	15	63	100	33	36	31	100	69	31	100	43	57	100
STATURE														
Tall ..	13	10	20	43	29	10	4	43	30	13	43	43	0	43
Short ..	9	5	43	57	4	26	27	57	39	18	57	0	57	57
Total ..	22	15	63	100	33	36	31	100	69	31	100	43	57	100

[1] By the general population I mean the population of England and Wales (the means were taken from the Anthropometric Committee of the British Association, *Brit. Ass. Ann. Report*, 1883, pp. 256f.). The definitions of the set of traits originally used also, for the most part, followed those of the same Committee. The original classification of hair- and eye-colour was somewhat more elaborate than that shown in the present tables (cf. 11, p. 150, 12 and refs.); but the sample is too small to permit of numerous fine subdivisions.

The reason for selecting the four characteristics set out in Table II will be obvious: as Dr. Morant has observed, "the classical anthropological classifications of the peoples of Europe were based almost entirely on the cephalic index, stature, and hair- and eye-colour; . . . it is not to be doubted that they make clear distinctions between the major groups in the continent as a whole." (*Races of Central Europe*, 1939, p. 135.)

We begin by dividing first the rows of Table II, and then the resulting columns, by the square root of the number in the corresponding diagonal cell (the number of persons characterized by each of the eight attributes). This converts the number in each diagonal cell to unity. The resulting figures are shown in Table III, which corresponds to the matrix R_t, p. 172. This is the matrix that we shall now proceed to factorize.

TABLE III. RELATIVE FREQUENCIES

Trait	F	R	D	L	M	B	N	W	T	S
HAIR										
Fair	1·000	·000	·000	·519	·213	·077	·359	·306	·423	·254
Red	·000	1·000	·000	·359	·215	·093	·342	·185	·394	·171
Dark ..	·000	·000	1·000	·241	·525	·611	·667	·430	·384	·718
EYES										
Light ..	·519	·359	·241	1·000	·000	·000	·566	·188	·770	·092
Mixed ..	·213	·215	·525	·000	1·000	·000	·401	·479	·254	·574
Brown ..	·077	·093	·611	·000	·000	1·000	·476	·290	·110	·642
HEAD										
Narrow ..	·359	·342	·667	·566	·401	·476	1·000	·000	·551	·622
Wide ..	·306	·185	·430	·188	·479	·290	·000	1·000	·356	·428
STATURE										
Tall	·423	·394	·384	·770	·254	·110	·551	·356	1·000	·000
Short ..	·254	·171	·718	·092	·574	·642	·622	·428	·000	1·000

Factor Analysis of Relative Frequencies. The saturations for the first factor, it will be remembered, can be determined by simply dividing the square root of the number of persons characterized by each attribute (N_i) by the total number ($N = 100$). The values so calculated are set out in the first column of Table VIIA. As may easily be verified, the factor-variance, i.e., the sum of the squares of these values, is $m = 4$. The reader may find it instructive to check the fact that the saturations correspond to the first principal component, by multiplying each row of Table III (or, say, the first two or three figures from each row) by the corresponding saturation, and dividing each weighted sum by the factor-variance, namely, 4. The results should be equal to the factor-saturations employed.

A hierarchy is now constructed by multiplying each saturation in turn by all the saturations. This is subtracted from Table III; and the residuals so obtained are given in Table IV.

TABLE IV. FIRST RESIDUALS

Trait	F	R	D	L	M	B	N	W	T	S
HAIR										
Fair ..	·780	−·182	−·372	·250	−·068	−·185	−·030	·045	·115	−·100
Red	−·182	·850	−·370	·137	−·017	−·123	·020	−·030	·140	−·121
Dark ..	−·372	−·370	·370	−·215	·049	·169	·008	−·012	−·136	·118
EYES										
Light ..	·250	·137	−·215	·670	−·345	−·320	·089	−·132	·393	−·341
Mixed ..	−·068	−·017	·049	−·345	·640	−·334	−·097	·145	−·139	·121
Brown ..	−·185	−·123	·169	−·320	−·334	·690	·013	−·020	−·255	·222
HEAD										
Narrow ..	−·030	·020	·008	·089	−·097	·013	·310	−·462	·006	−·005
Wide ..	·045	−·030	−·012	−·132	·145	−·020	−·462	·690	−·009	·008
STATURE										
Tall	·115	·140	−·136	·393	−·139	−·255	·006	−·009	·570	−·495
Short ..	−·100	−·121	·118	−·341	·121	·222	−·005	·008	−·495	·430

A minor peculiarity in the three foregoing tables deserves a word of comment. In the diagonal submatrices of Tables II and III, all the values except those in the leading diagonal are zero, and thus "all the frequency is concentrated in the diagonal compartments." [1] This is plainly due to the fact that all the alternative determinates under one and the same determinable are, by hypothesis, mutually exclusive: that being so, of the 22 persons who are entered as having fair hair none can also be entered as having red or dark hair. In the corresponding cells of Table IV the negative residuals indicate that, after allowing for the effect of the first or general factor, we find that any tendency to grow fair hair inevitably has a negative effect on the tendency to grow hair of any other colour.

The next step is to test the residuals for significance. For this purpose we must obviously confine our examination to those residuals resulting from figures that have been actually observed, i.e., those appearing in the non-diagonal matrices, relating to *different* determinables. If the reader calculates the sum of the squares of the figures in each of the side submatrices in Table IV, and then multiplies that sum by $N = 100$, he will find that the results agree exactly with the figures obtained for chi-squared when calculated in the usual way from the corresponding submatrices in Table I. The six values thus obtained are shown in Table V. All except those for the association of head-shape with hair-colour and stature are significant. The total value for chi-squared comes to 78·35, which, with 13 degrees of freedom, is highly significant.

TABLE V. CHI-SQUARE

Determinable	HAIR	EYES	HEAD	STATURE
HAIR ..	—	21·26	0·45	9·01
EYES ..	21·26	—	5·63	41·98
HEAD ..	0·45	5·63	—	·02
STATURE	9·01	41·98	·02	—

TABLE VI. POINT-CORRELATIONS

Determinable	HAIR*	EYES	HEAD	STATURE
HAIR ..	1·000	·386	−·043	·284
EYES ..	·386	1·000	·195	·553
HEAD ..	−·043	·195	1·000	·014
STATURE	·284	·553	·014	1·000

The residuals in Table IV may be regarded as expressing partial or specific associations. They can be factorized in their turn; and the process repeated. Since there are only seven degrees of freedom, only seven factors can be extracted. All except the last are fully significant. The saturations [2] are given in Table VIIA. Most of the factors can readily be interpreted. The first is a general factor, i.e., all its signs are positive: as we have already seen, it specifies the general pattern of traits (or trait-tendencies) which characterizes the particular sample of the British population we have been examining: in the main they tend to be dark-haired, of mixed eye-colour, dolichocephalic, and short rather more frequently than tall. The second factor is bipolar: it indicates a contrast between (*a*) a complex tendency which combines light eyes with tall stature, fair or red hair, and a narrow head, and (*b*) an opposite tendency to combine short stature with dark hair, brown or (less frequently) mixed eye-colour, and a broad head: of the two antithetical types light eyes in the one case and short stature in the other appear to be the traits which are most diagnostic. The third and fifth factors appear to contrast a brachycephalic tendency (and its associated characteristics) with a dolichocephalic; the fourth to contrast the red-haired type with the fair-haired; and the sixth suggests a minor type combining tall stature with dark hair and light or mixed eye-colour.

I do not suggest that the particular analysis given is either a final analysis or the best analysis. For one thing, several of the factor-saturations, e.g., those for fair and red hair in factors III and IV, suggest that (if only a slightly larger number was available) a group factor analysis or a method of subdivided factors would yield a clearer picture. Further, the

[1] The description is quoted from Yule, *loc. cit.*, p. 65.

[2] The factor-saturations have been calculated by the method of weighted summation. They therefore represent 'principal components.' Accordingly, the reader, if he wishes, can apply Bartlett's test for significance (cf. this *Journal*, III, pp. 78, 118). But in cases like the present I should consider the chi-squared test more appropriate, since this can be calculated for particular submatrices only.

I should add that I am deeply indebted to Mr. Arthur Summerfield for assistance with the computations.

large saturations (over 0·6) that stand in isolation—for eye-colour in factor II, head-shape in factor III, and hair-colour in factor IV—tempt us to ask whether we might not secure a truer result if, instead of retaining 1·000 in the leading diagonal, we substituted a reduced figure which would exclude specific factors. But to discuss these further issues would take us too far from our immediate task.

TABLE VII. BIPOLAR FACTOR SATURATIONS

A. Analysis of Relative Frequencies								*B. Analysis of Point-Correlations*				
Traits	Factors							Traits	Factors			
	I	II	III	IV	V	VI	VII		A	B	C	D
HAIR												
Fair	·469	·420	·231	−·602	−·100	−·404	·119					
Red	·387	·333	−·019	·751	·182	−·376	·012	HAIR	·625	·508	·429	−·410
Dark	·794	−·411	−·128	−·011	−·030	·422	−·076					
EYES												
Light	·574	·719	−·164	−·104	·042	·084	−·325					
Mixed	·600	−·241	·563	·227	−·442	·048	·125	EYES	·820	−·176	−·478	−·410
Brown	·557	−·482	−·437	−·138	·443	−·138	·201					
HEAD												
Narrow	·831	·081	−·428	·039	−·338	−·044	·056	HEAD	·448	−·709	·478	·410
Wide	·557	−·121	·638	−·058	·504	·065	−·084					
STATURE												
Tall	·656	·619	·010	·047	·120	·335	·241	STATURE	·711	·378	−·429	·410
Short	·755	−·537	−·009	−·041	−·104	−·292	−·209					

Factor Analysis of Point-Correlations. As we have already seen, in dealing with qualitative attributes of a dichotomous type, i.e., in cases where each determinable furnishes only *two* mutually exclusive determinates, a set of product-moment correlations derived from fourfold tables will commonly provide the simplest basis for factor analysis. But it is, I think, easy to show that for data of the present type, where several of the determinables lead to a multiple classification, this method is far from satisfactory. However, let us briefly examine its application here.

In Table II, one of the submatrices (that for Head and Stature) involves a single degree of freedom only. For this pair of variables we can at once combine the four figures there given into a single figure expressing the root-mean-square contingency ($\varphi = \sqrt{\chi^2/N}$). This may be regarded as giving the product-moment correlation for the two determinables on the assumption that the frequencies for the four determinates have a point-distribution. Where the submatrices contain more than four figures, we can, if we wish, calculate a coefficient of contingency by Pearson's formula, or better, allow for the increased degrees of freedom, by dividing the corresponding φ by the fourth root of the degrees of freedom.[1] But it is better still to pool the initial data so as to secure a fourfold table for every pair of determinables from the outset. In view of the results just obtained we may plausibly group fair and red hair together ; for eye-colour the original records make it possible to divide most of those classed above as ' mixed ' either with the ' light ' or with the ' brown eyes ' respectively, and the rest may be divided in due proportion.[2] We can then calculate point-correlations for the resulting fourfold tables. In doing so, I have taken what we may call Nordic traits (fair hair, light eyes, narrow head, and tall stature) to be positive qualities. The coefficients thus obtained are shown in Table VI.

[1] This expression has been proposed by A. A. Tschuprow (8). It appears to have been suggested by the ' divergence coefficient ' of Lexis (1) : see also 13, p. 325. Here the coefficients so calculated are in moderately close agreement with the φ-coefficients set out in Table VI.

[2] It might perhaps be argued that we could preserve the distinction between fair hair and red hair by constructing *two* sets of fourfold tables for hair-colour, based on the dichotomies (i) ' fair ' or ' not-fair (i.e., red and dark) ' and (ii) ' red ' or ' not-red (i.e., fair and dark).' But both logical considerations and the empirical results make it clear that this alternative procedure only increases the confusion when we come to interpret the factor-saturations so obtained.

These coefficients can readily be factorized ; and the saturations, calculated by simple summation, are shown in Table VIIB. It will be seen at once that, without the more detailed knowledge reached by the method previously described, it would be very difficult to arrive at a convincing interpretation. The first factor (A) obtained from the table of point-correlations obviously corresponds, as we should expect, with the second factor (II) obtained from the table of relative frequencies : thus factor A expresses as a general factor what factor II expressed as a bipolar factor, namely, the tendency of the four positive (Nordic) traits to go together. Factor D expresses the contrast between (*a*) the combination of dark hair and dark eyes and (*b*) wide head and short stature shown by the saturations for these qualities in factor III. Factor C expresses the contrast between (*a*) the combination of dark hair and wide head and (*b*) the combination of brown eyes and short stature suggested by factor IV. Factor B apparently expresses the contrast between (*a*) the combination of red hair and tall stature and (*c*) light eyes and narrow head expressed by factor IV. However, with only four determinables, and with unity in the leading diagonal, it might well be maintained that the last three factors mainly represent specific factors expressed in bipolar form (see 10, pp. 112f.).

It must therefore be admitted that cases of manifold classification—i.e., determinables involving more than two determinates—cannot as a rule be satisfactorily treated by forcing the multiple determinates into twofold form and then analysing the coefficients derived from the resulting fourfold tables.

TABLE VIII. GROUP FACTOR SATURATIONS

Analysis of Point Correlations

Traits	Factors			
	A	B	C	D
HAIR ..	−·043	·999	·000	·000
EYES ..	·195	·395	·457	·773
HEAD ..	1·000	·000	·000	·000
STATURE	·014	·285	·958	·000

It is always instructive to supplement a bipolar analysis by a group factor analysis. The patterns exhibited by Tables VI and VIIB, suggest that the most appropriate procedure will be that which I have described elsewhere for cases in which the variables reveal a ' cumulative complexity.' [1] The result is a matrix of ' completely overlapping ' (as distinct from ' partly overlapping ' or ' non-overlapping ') group factors (Table VIII). Here the implication of the figures is that the first classification should be based predominantly on head-shape ; the second on hair-colour, eye-colour, and stature, with greatest emphasis on hair-colour ; the third on the joint variation of eye-colour and stature, regardless of hair-colour ; and the last on divergences in eye-colour from that indicated by the preceding classifications.

The Examination of Ultimate Frequencies. Had we been factorizing a correlation table of the ordinary kind for quantitative variables, we should, after making a bipolar analysis like that shown in Table VIIA, have proceeded at once to take the bipolar factors as indicating the way in which the several traits might be combined into groups, and then use this classification as the basis for the analysis into group factors. With qualitative traits a more direct procedure is available for determining what groupings are most significant.

Following Yule's terminology, and modifying his definition slightly so as to include those cases where the classifications may be manifold, let us term the classes specified by *all* the determinates that could be observed in any case, i.e., the classes of the *m*th order in the case of *m* determinables, the ' ultimate frequencies.' [2] As Yule points out, " the ultimate frequencies form a natural set in terms

[1] *Factors of the Mind*, p. 307. As there pointed out, the factor-matrix so obtained has the form of a triangular matrix, as will here be evident if the traits are re-arranged in the order described in the text (Head, Hair, Stature, Eyes). The implications suggested are illustrative only : to have any genuine value they should be based on a far larger number of traits.

[2] The definition, as phrased by Yule (6, p. 11), specifies ' attributes ' only ; he does not distinguish attributes which are determinates from those which are determinables. Such a distinction is not required by the types of problem with which he chiefly deals, namely, dichotomous classes only (presence or absence) in the case of many determinables, and two determinables only in the case of manifold determinates.

of which the data are *completely* given " (*ibid.*, p. 13). With a large number of determinates (such as may arise in certain types of test) it would be exceedingly laborious to examine all the ultimate frequencies in turn. However, in such cases the preliminary bipolar analysis would help us to pick out those sets of traits which go together most closely and are likely to form significant *m*th order combinations. In the present case both *m* and the n_i's are small numbers ; and consequently the *total* number of *m*th order combinations, namely, $n_1 \times n_2 \times n_3 \times n_4 = 3 \times 3 \times 2 \times 2$ amounts to 36 only. Here therefore it will be as simple as it is instructive to set out *all* the 4th order combinations, and examine the significance of each one.

In Table IX (last column but one) I give the frequencies for each of the 36 composite categories. The number expected by chance is shown in the last column. The probability

TABLE IX. THE DETERMINATION OF SIGNIFICANT GROUPS

Hair	Eyes	Head	Stature	Ultimate Frequencies Obs.	Exp.
Fair	Light	Narrow	Tall	8	2·1
,,	,,	,,	Short	4	2·8
,,	,,	Wide	Tall	2	1·0
,,	,,	,,	Short	0	1·3
	Total			14	7·3
,,	Mixed	Narrow	Tall	1	2·3
,,	,,	,,	Short	1	3·1
,,	,,	Wide	Tall	2	1·1
,,	,,	,,	Short	2	1·4
	Total			6	7·9
,,	Brown	Narrow	Tall	0	2·0
,,	,,	,,	Short	0	2·7
,,	,,	Wide	Tall	0	0·9
,,	,,	,,	Short	2	1·2
	Total			2	6·8
Total				22	22·0
Red	Light	Narrow	Tall	6	1·5
,,	,,	,,	Short	0	1·9
,,	,,	Wide	Tall	2	0·7
,,	,,	,,	Short	0	0·9
	Total			8	5·0
,,	Mixed	Narrow	Tall	2	1·6
,,	,,	,,	Short	1	2·1
,,	,,	Wide	Tall	0	0·7
,,	,,	,,	Short	2	1·0
	Total			5	5·4
,,	Brown	Narrow	Tall	0	1·4
,,	,,	,,	Short	2	1·8
,,	,,	Wide	Tall	0	0·6
,,	,,	,,	Short	0	0·8
	Total			2	4·6
Total				15	15·0

TABLE IX (*continued*).

HAIR	EYES	HEAD	STATURE	ULTIMATE FREQUENCIES Obs.	Exp.
Dark	Light	Narrow	Tall	9	6·1
,,	,,	,,	Short	0	8·2
,,	,,	Wide	Tall	2	2·8
,,	,,	,,	Short	0	3·7
	Total	.. ..		11	20·8
,,	Mixed	Narrow	Tall	3	6·7
,,	,,	,,	Short	12	8·9
,,	,,	Wide	Tall	2	3·1
,,	,,	,,	Short	8	4·0
	Total	.. ..		25	22·7
,,	Brown	Narrow	Tall	1	5·8
,,	,,	,,	Short	19	7·7
,,	,,	Wide	Tall	3	2·6
,,	,,	,,	Short	4	3·4
	Total	.. ..		27	19·5
TOTAL	.. ..	.. ..		63	63·0
GRAND TOTAL	.. ..	.. ..		100	100·0

that a given person will be at once fair-haired, light-eyed, dolichocephalic, and tall will be $\frac{22}{100} \times \frac{33}{100} \times \frac{69}{100} \times \frac{43}{100} = \cdot 0215$. Accordingly, in a sample of 100 we should anticipate that only 2·15 persons would exhibit that combination. Actually there are as many as 8. As the theoretical proportions are small we can take the standard error of the discrepancy to be approximately equal to the square root of the expected number, namely $\sqrt{2 \cdot 15}$. Hence in this case the excess over expectation is unquestionably significant.

A glance down the two columns will show that the following combinations may be regarded as fully significant :

(i) Fair hair, light eyes, narrow skull, tall stature.
(ii) Red hair, light eyes, narrow skull, tall stature.
(iii) Dark hair, brown eyes, narrow skull, short stature.

The following fall somewhat below the level of significance in the present group, but were fully significant in the larger group of 217 :

(iv) Dark hair, eyes of mixed colour, broad skull, short stature.
(v) Dark hair, light eyes, narrow skull, tall stature.

In the larger group one further combination nearly reached the level of significance, namely :

(vi) Fair hair, light or mixed eyes, broad skull, tall stature.

The first two combinations are suggestive of two familiar varieties of the so-called Nordic type ; the third is suggestive of the so-called Mediterranean type ; the fourth of the so-called Alpine or Celtic type ; the fifth is a recognized subtype usually supposed to result from a crossing of Nordic and Mediterranean types ; the sixth is suggestive of the so-called Beaker-folk. I may add that it was noted at the time that persons belonging to the first two groups frequently came from areas such as the Lake District and Yorkshire, where tall, fair-haired, light-eyed populations predominate ; however, a number of the red-haired members came from Ireland (chiefly north-eastern areas). The third group included a large proportion of Welsh or Irish immigrants ; the fourth group was predominantly Welsh, and the fifth (a very small group) predominantly Irish ; the sixth group also came

chiefly from Wales. In my original analysis, place of origin was also included as a fifth determinable ; but the numbers in the resulting subgroups were too small to yield significant figures. Migration and mixed marriages likewise made it clear that, with the data available, a satisfactory classification in regard to ancestral origin, which alone is really relevant, was scarcely possible.

If we regard each significant combination of traits as resting on a factor, it will be seen that in the main the results thus reached restate, in group factor form, the classification already arrived at by the bipolar analysis. It is obvious that we could pass from one type of analysis to the other by rotation ; and this would yield appropriate weights for the group factors if required.

Finally, let me say that in this short paper I have only been able to touch very briefly on a few of the points raised by the problem I have been discussing and to sketch, in the barest outline, one of the many possible solutions.

V. SUMMARY AND CONCLUSIONS

1. It is argued that, if factor analysis is to be adequately employed in problems of individual psychology, a rigorous statistical treatment is needed for qualitative data no less than for quantitative ; and it is claimed that, with minor modifications, factorial methods can be applied, as legitimately, as readily, and as fruitfully, with the former as with the latter.

2. The various techniques available for examining qualitative data are enumerated and briefly discussed. It is concluded that, with dichotomous classifications that are assumed to rest on a discrete point-distribution, the product-moment correlation for point-distributions (φ) will, in general, prove the most useful.

3. Cases of manifold classification, in which there are many determinables, each involving many mutually exclusive determinates, have been almost entirely ignored by previous writers. For these cases, it is shown, the point-distribution coefficient, as ordinarily calculated, is unsuitable.

4. The general methods here proposed for such cases rest on the assumption that factor analysis can be validly adopted with any type of symmetrical positive-definite matrix. The procedure suggested as appropriate for most purposes proceeds by analysing the relative frequencies. This leads to a factor matrix containing one general factor and a succession of bipolar factors. The theoretical number of factors will, as usual, depend on the rank of the matrix analysed, which will always be less than the total number of attributes. It is shown that this mode of analysis permits of employing the chi-squared criterion for testing the significance to the residuals obtained.

5. The group factor method also can be applied direct to the table of relative frequencies. But more instructive results can be obtained by examining the ultimate frequencies, or (with large tables) that particular selection of ultimate frequencies which the bipolar analysis indicates is likely to be most significant.

6. The procedure is illustrated by an analysis of the physical characteristics commonly adopted in anthropological studies of physical types, namely, hair- and eye-colour, head-shape and stature. The factors, and significant groupings found, agree in the main with those combinations alleged to be characteristic of the commoner physical types described in anthropological surveys of the British Isles. It is claimed that a factorial study of the way in which such traits are associated or combined is likely to lead to more precise and trustworthy results than the usual method of reporting averages for the separate traits.

REFERENCES

1. Lexis, W. (1886). 'Uber die Theorie der Stabilität Statischer Reihen,' *Jahrb. f. nat. Ök. u. Stat.* XXVII, 604–48
2. Yule, G. U. (1900). 'On the association of attributes in statistics,' *Phil. Trans.*, A, CXCIV, 257–84.
3. Pearson, K. (1900). 'On the correlation of characters not quantitatively measurable,' *Phil. Trans.*, A, CXCV, 1–87.
4. Pearson, K. (1904). 'On the theory of contingency and its relation to association and normal correlation,' *Drapers' Company Research Memoirs, Biometric Series*, I.
5. Lipps, G. F. (1905). 'Die Bestimmung der Abhängigkeit Zwischen den Merkmalen eines Gegenstandes,' *Ber. d. Math. phys. Klasse d. Kön. Sächs. Gesellsch. d. Wissenschaften.*
6. Yule, G. U. (1912). *Introduction to the Theory of Statistics.* 2nd ed. London : Griffin.
7. Burt, C. (1912). 'The inheritance of mental characters,' *Eugen. Rev.*, IV.
8. Tschuprow, A. A. (1918–9). 'Zur Theorie der Stabilität Statistischer Reihe,' I, 199–256, II, 80–133.
9. Burt, C. (1921). *Mental and Scholastic Tests.* London : P. S. King (2nd ed., 1946).
10. Burt, C. (1944). 'The factorial study of physical types,' *Man*, XLIV, 82–6.
11. Burt, C. (1945). 'The relation between eye-colour and defective vision,' *Eugen. Rev.*, XXXVII, 149–56.
12. Burt, C., and Grieve, J. (1945). 'Defective colour vision in relation to pigmentation and hair,' *Man*, XLV, 81–3.
13. Kelley, T. L. (1947). *Fundamentals of Statistics.* Cambridge, Mass. : Harvard University Press.
14. Burt, C. (1949). 'Alternative methods of factor analysis,' *Brit. J. Psychol., Stat. Sect.*, II, 98–121.

A NOTE ON SHELDON'S TABLE OF CORRELATIONS BETWEEN TEMPERAMENTAL TRAITS

By A. LUBIN
University College and Maudsley Hospital, London

I. *Problem.* II. *Criterion.* III. *Results.* IV. *Summary.*

I. PROBLEM

In a recent paper (1) Adcock has discussed the table of correlations between temperamental traits published by Sheldon in his book on *The Varieties of Temperament* (6), and reports considerable difficulty in finding a method of factor analysis which will account for the table as it stands. The correlations are reproduced below in Table I. As Professor Burt has pointed out, the symmetrical pattern of positive and negative signs is unlike that found in any other published table. Adcock tried various methods of factor analysis. In each case it proved impossible to complete the analysis since it led to illegitimate procedures such as taking the square root of a negative number. His own suggestion was that, to explain Sheldon's three basic variables, a complex interaction between *four* factors would have to be postulated. " The four factors " (described in the closing paragraphs of his article) " seem to lend themselves to precisely the sort of interaction which we suspected, and seem capable of explaining the unusual form of the correlations."

There is, however, a simpler hypothesis. Earlier in his paper Adcock observes : " obviously there is something peculiar about these inter-correlations." With this remark we may readily agree. And the peculiarity is so great that one is forced to ask whether it may not be outside the bounds of mathematical possibility. Hence the important question to ask is not : can these correlations be factorized ? but rather the prior question : are these correlations mutually compatible ?

In the following note, it is proposed to demonstrate that at least three of Sheldon's product-moment correlations could not be obtained simultaneously from any actual set of measurements : in short, that the table violates the well-known conditions for consistency. It follows that some at least of his figures must contain errors of arithmetical calculation.

II. CRITERION

Such errors can frequently be detected in correlation tables in virtue of certain inevitable boundary conditions that must obtain for any set of inter-correlations of three or more variables. Yule in an early paper points out that, if r_{12} and r_{13} are known, definite limits can be laid down for r_{23} : (8, p. 250 and refs. ; cf. 3, p. 387, and 5, p. 142). These may be stated in terms of r_{12} and r_{13} as follows :

$$r_{12}\,r_{13} - \sqrt{1 - r_{12}^2}\,\sqrt{1 - r_{13}^2} \leqslant r_{23} \leqslant r_{12}\,r_{13} + \sqrt{1 - r_{12}^2}\,\sqrt{1 - r_{13}^2} \,. \quad \text{(i)}$$

Yule shows that the foregoing equation results from the fact that the partial correlation, $r_{23\cdot1}$, cannot exceed $+1$ and cannot be less than -1. This is true for any set of product-moment correlations (cf. 3, p. 330, and 8, pp. 238f.). The same conditions, as Burt has pointed out, can also be derived from the fact that the squared multiple correlation cannot exceed unity. Yule goes on to prove that analogous conditions obtain for the consistency of the correlations of between four or more variables : this can readily be deduced from the fact that $r_{12\cdot34}$, $r_{12\cdot345}$, etc., must lie between $+1$ and -1.

TABLE I. CORRELATIONS BETWEEN TRAITS

Trait		*V*10	*V*14	*V*8	*V*9	*S*3	*S*5	*S*6	*S*10	*C*8	*C*1	*C*6	*C*5
Greed for affection	*V*10	–	·81	·82	·79	−·39	−·33	−·38	−·31	−·45	−·35	−·40	−·36
Complacency	*V*14	·81	–	·75	·76	−·42	−·36	−·35	−·33	−·38	−·44	−·41	−·45
Sociophilia	*V*8	·82	·75	–	·79	−·30	−·37	−·41	−·38	−·52	−·44	−·35	−·37
Amiability	*V*9	·79	·76	·79	–	−·40	−·36	−·37	−·32	−·44	−·39	−·45	−·35
Energetic characteristic	*S*3	−·39	−·42	−·30	−·40	–	·74	·81	·75	−·49	−·71	−·60	−·59
Love domination	*S*5	−·33	−·36	−·37	−·36	·74	–	·85	·78	−·67	−·58	−·67	−·56
Love risk and chance	*S*6	−·38	−·35	−·41	−·37	·81	·85	–	·82	−·68	−·69	−·59	−·59
Psychiatric callousness	*S*10	−·31	−·33	−·38	−·32	·75	·78	·82	–	−·67	−·63	−·60	−·69
Sociophobia	*C*8	−·45	−·38	−·52	−·44	−·49	−·67	−·68	−·67	–	·77	·84	·79
Restraint posture and movement	*C*1	−·35	−·44	−·44	−·39	−·71	−·58	−·69	−·63	·77	–	·79	·82
Emotional restraint	*C*6	−·40	−·41	−·35	−·45	−·60	−·67	−·59	−·60	·84	·79	–	·82
Apprehensiveness	*C*5	−·36	−·45	−·37	−·35	−·59	−·56	−·59	−·69	·79	·82	·82	–

So long as we are concerned solely with the consistency of Sheldon's figures, and not with their factorial analysis, we can disregard the *positive* correlations within each of the three main groups, *V*, *S*, and *C*, and scrutinize only the negative correlations between the different groups. In verbal form the problem might be stated thus : we can say that *two* things are the opposites each of the other ; when (if ever) can we say that *three* things are opposites ? The answer evidently is : only when the opposition is more or less incomplete. If opposition is expressed in quantitative form by a negative correlation, then (as the formula shows) three negative coefficients, all numerically larger than −0·50, would be impossible. Accordingly, in Table I any negative values well over −0·60 are likely to be highly suspect.

To obtain a decisive test, partial correlations involving three variables at a time were computed for each of the triads in Sheldon's correlation table that appeared suspicious. The well-known Pearson-Yule formula for partial correlation was employed, namely,

$$r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} \quad . \quad . \quad . \quad .(ii)$$

III. RESULTS

Out of nineteen partial correlations computed in this manner, as many as five furnished negative coefficients numerically larger than −1. The detailed figures are set out below in Table II.

TABLE II. PARTIAL CORRELATIONS SHOWING IMPOSSIBLE VALUES

Correlated Traits	Trait Eliminated	Partial Correlations
*V*8, *S*5	*C*8	−1·13
*V*8, *S*6	*C*1	−1·09
*V*8, *S*6	*C*8	−1·22
*V*9, *S*3	*C*1	−1·04
*V*14, *S*3	*C*1	−1·16

It is evident that, if only one triad were found to give an impossible partial correlation, then at least one of the three first-order correlations would be in error. In the present case, the analysis suggests that, at the very least, three first-order correlations must involve errors of computation. With the data available it is impossible to determine with certainty which coefficients are in error and which may be safely accepted. The large negative values printed by Sheldon for the correlations between *S*3 and *C*1 (−0·71), *S*6 and *C*1 (−0·69), and *V*8 and *C*8 (−0·51) would account for all the anomalies. Or again, since the variables *V*8 and *C*1 occur three times, and one or other of the two appears in each of the five cases, mistakes over these two would suffice.

It will be noted that no question of linearity, sampling error, normality, homogeneous variance, etc., is involved. The matter is simply one of arithmetical accuracy.

The results emphasize the supreme importance of using methods of computation which include a complete system of checks (4, 7). A recent letter by Greenall in the *Journal* (2) makes it plain that it is not only tables of inter-correlations that can be inaccurate : errors can of course slip through even the most complex set of checks, but it is unlikely that large or numerous errors will result if one of the standard checking methods is adopted. Such methods should be treated as a matter of routine with all data intended for publication.

Finally, I wish to thank Mr. A. R. Jonckheere for assistance in the calculations made for this paper, and Professor Burt for guidance on several points, including the reference to Yule.

IV. SUMMARY

1. Difficulties have been experienced in factorizing Sheldon's table of correlations between temperamental traits. It is suggested that this is due to peculiarities, not in the underlying factors, but rather in the manner in which the correlations have been obtained : the pattern of negative coefficients is quite exceptional, and some of the coefficients appear to be arithmetically impossible.

2. This is confirmed by applying Yule's criterion,[1] namely, that, in an empirical table, no partial correlations can exceed unity. At least three of Sheldon's coefficients appear to be erroneously computed.

3. As a corollary it is suggested that all correlations should, as a matter of routine, be calculated by some method which permits of arithmetical checks.

REFERENCES

1. Adcock, C. J. (1948). ' A factorial examination of Sheldon's types.' *Journ. Pers.*, XVI, 312–19.
2. Greenall, P. D. (1949). ' Two criticisms.' *Brit. J. Psychol., Stat. Sect.*, II, 64–5.
3. Kendall, M. G. (1947). *The Advanced Theory of Statistics.* Vol. I. London : Griffin.
4. Kossack, Carl F. (1948). ' On the computation of zero-order correlation coefficients.' *Psychometrika*, XIII, 91–3.
5. McNemar, Quinn (1949). *Psychological Statistics.* New York : Wiley.
6. Sheldon, W. H. (1942). *The Varieties of Temperament.* New York : Harper.
7. Tucker, L. R. (1948). ' A note on the computation of a table of intercorrelations.' *Psychometrika*, XIII, 245–50.
8. Yule, Udny (1912). *An Introduction to the Theory of Statistics.* London : Griffin.

[1] In this article my chief purpose has been to demonstrate the importance of checking the arithmetical accuracy of the figures appearing in any table of inter-correlations and to show that those in Sheldon's table are mutually inconsistent. Professor Burt, however, points out that with other tables the criterion given above in equation (i) might be satisfied perfectly, and yet the coefficients remain inconsistent. " For example," he writes (in a personal communication), " a very slight change in Sheldon's table would suffice to reduce the partial correlations *of the first order* to an acceptable size ; and yet, owing to the large negative correlations between Sheldon's *S*'s and *C*'s, it would still be impossible to analyse the table in terms of real factors." The more general criterion, that he has suggested elsewhere, consists in showing that *all the leading principal minors* are non-negative : when this has been proved, the matrix is positive definite (or semidefinite). These minors emerge automatically when the partial correlations are computed by what he calls the ' factorial method ' (cf. *Factors of the Mind*, p. 307, and *Laboratory Notes on Elementary Statistics*, 1941, sect. 8) : with this method therefore no very elaborate calculation would seem to be required.

I should like to take this opportunity of noting two errata in my previous article, this *Journal*, III, pp. 95 and 102. On p. 95 the plus sign under the radical should be minus ; on p. 102 in the formula for c, read $\sqrt{N}$ for N.

BOOK REVIEWS

Human Ability. By C. Spearman and Ll. Wynn Jones. Macmillan & Company. Pp. vii + 198. 16s. net.

This slender volume embodies Professor Spearman's last work. It has been long awaited, and is addressed to all psychologists whether they happen to be interested in the statistical approach or not. Written in collaboration with Dr. Wynn Jones, it provides a comprehensive summary of Professor Spearman's final views on the factorial study of the mind—its methods, controversies, and results, and their psychological implications. The chapters suffer a little from the attempt to compress so much material into so small a space ; and at times, in the effort to avoid excessive technicality, the writing drops into picturesque metaphor and loose colloquialisms, which often obscure the meaning instead of clarifying it. The manuscript was unfortunately left in an unfinished state ; and, partly no doubt as a result, the text is somewhat inadequately annotated, and the references are not always complete or exact.

The problem with which the writers are chiefly concerned is the place to be allotted to ' group factors.' Those who are familiar with recent literature will be somewhat disappointed with the outcome of the discussion, for the views expressed are rather confused, and show little advance beyond the position adopted in *The Abilities of Man* some twenty years ago. There, it will be remembered, Professor Spearman, while still defending his celebrated theory of two factors, at the same time tried to leave a loophole for ' exceptional cases,' should they be finally confirmed.

In the present volume, we are assured, " the psychological picture still remains essentially the same." Nevertheless, in the summary to chapter II (headed " Theory of Two Factors ") the writers tell us that " the statistical results of this method . . . consist of (1) a general factor ; (2) an unlimited number of narrow specific factors ; and (3) very few broad group factors." This formulation is almost identical, word for word, with the statement of the " three-factor theory," put forward by Burt over thirty years ago,[1] which Spearman at the time so strongly criticized. But how, the puzzled reader is tempted to enquire, can a two-factor method yield three kinds of factor ? And, even if " very few " group factors are to be accepted, this nevertheless means that the two-factor theory, which postulates general and specific factors only, is no longer adequate.

Elsewhere, it is true, the authors try to set up a distinction between the kind of supplementary factors they are prepared to admit and the ' group factors ' or ' special abilities ' of earlier writers. Group factors, they say, must always depend on the tests put into the battery, and consequently cannot represent innate capacities. Their own factors they prefer to describe as ' broad *specific* factors ' which, we are told, " must be sharply distinguished from what have been called *group* factors," although, a little inconsistently, they are termed ' group factors ' in chapter II.

A new feature in their discussion is the development of a ' Method of Excess Correlation,' which is put forward as an alternative to existing group-factor methods. To illustrate the procedure, the writers print an artificial table of correlations arranged in such a fashion that all the highest correlations appear in the diagonal submatrices. This, of course, is the way in which an empirical correlation table is usually arranged before Burt's group-factor method is applied ; and, as he has himself pointed out, the underlying principle goes back to Karl Pearson, who distinguished, in tables of correlations between physical measurements, two different levels of correlation, namely, the high positive correlations between ' like ' traits and the comparatively low positive correlations between ' unlike ' traits respectively. The general scheme had already been set out and criticized by Spearman and Hart in their well-known paper (*Brit. J. Psych.*, V, 1913, p. 57), where they claimed to show that it is *not* encountered in correlations between mental measurements. On this point, therefore, Spearman seems plainly to have changed his view. As regards the new method here proposed, an obvious defect arises from the fact that the formula necessarily assumes that all the cross-correlations, in which only the general factor is operative, are equal in numerical value.

[1] E.g., in *The Distribution and Relations of Educational Abilities* (1917) and the *Report on Psychological Tests of Educable Capacity* (1924).

Such an assumption is rarely fulfilled, even approximately, by the figures that are found in empirical tables.

Here and elsewhere neither the basic part played by Pearson in developing the product-moment formulæ, the principles of multiple correlation, and the method of principal axes, nor yet his many constructive criticisms, are ever explicitly mentioned. In an early chapter on " Some Criticisms," and later in the text, objections raised by Pearson and his followers are certainly cited, and refutations are offered ; but Pearson himself is never referred to by name. This ban (as older psychologists in this country are well aware) dates from a heated controversy between the two which arose as far back as 1906. But by now it could surely be lifted. These anonymous references, to say the least, are highly misleading to the younger student.

To the statistical reader the most interesting section will doubtless be that which reviews and criticizes the methods of factor analysis developed by more recent workers. Spearman naturally takes first the ' bifactor method ' put forward by Holzinger, since that claims to be based on the principles underlying his own two-factor theory : the method he accepts as valid in " certain conditions where the Two Factor analysis ceases to be applicable at all." He turns next to ' simple and weighted summation.' Here he makes quite clear his main objections to Burt's procedures. " Chronologically even earlier," he writes, " came another outstanding attempt at extending the scope of factorizing ; this was contributed by Cyril Burt : whereas the older method dealt with every tetrad separately, the newer way took the sum of each whole column of correlations." All such attempts, however, he considers to be unsuccessful. His chief objection is to " the step whereby the vacant diagonal cells are filled out (by stopgaps) so as to make the whole table into a solid rectangular block." Thirdly, he briefly examines Thurstone's centroid method, and quotes his ' fundamental equation.' " This," he says, " is Burt's formula " ; hence much the same criticisms are urged. And finally he glances at the method of ' principal components,' in which " the axes of the (correlation) ellipsoids correspond to the factors." This is attributed to Kelley and Hotelling, and no reference is made to the earlier outline of the procedure developed by Karl Pearson. It is summarily dismissed on the ground that, like the centroid method, it produces negative as well as positive saturations in the supplementary factors. Burt's bipolar factors Spearman appears to accept, at any rate in the temperamental field (p. 60). Up to now, he says, unlike the group and general factors, " they have attracted comparatively scant attention : but recently they have come more and more into their proper estate as research is turning to the other side of human nature—feeling and volition." In the cognitive field he rejects them, arguing that intellectual abilities cannot have a negative influence on actual performance. This criticism is plainly due to the fact that for Spearman a factor is always a causal factor, never a principle of classification. Consequently, he fails to understand that, in the view of the writers he is criticizing, when (say) a non-verbal test receives a negative saturation, what is negatived is the verbality, not its influence on actual performance. With Holzinger, he tells us, " the concept of ' bipolar factor ' has been greatly extended " ; and he thinks that " in this sphere of factor we must look to him for the fullest development."

What he calls ' indirect methods,' such as ' rotation ' and the use of ' oblique axes,' are rejected because they too readily permit a subjective choice of whatever factors are required by the investigator's personal theories. His own method, he says, has been throughout that of partial correlation, with an *external* criterion rather than an internal. Taking external criteria (such as teachers' assessments for *g* or the Binet tests of intelligence) the general factor is partialled out by Yule's formula, and the residuum examined for supplementary factors. This, however, is scarcely factor analysis as now understood, since that proceeds primarily by seeking *internal* criteria derived from the table of inter-correlations themselves.

To those who knew Professor Spearman, it will be clear that he had lost none of his statistical ingenuity and little of his fighting spirit. To stress too strongly the shortcomings of the book would perhaps be unfair. As a pupil and devoted follower of Professor Spearman, Dr. Wynn Jones has evidently had a difficult task in preparing an incomplete manuscript for the press ; and the labour which he has expended on the work shows how loyally he has carried it out. Everyone will feel deeply indebted to him for the way in which he has endeavoured to make Professor Spearman's final views available to the psychological reader. *E. K. C. B.*

Psychological Statistics. By Quinn McNemar. New York : John Wiley & Sons, Inc. London : Chapman & Hall. 1949. Pp. viii + 364. $4.50.

Dr. McNemar is Professor of Psychology, Statistics, and Education at Stanford University ; and his text-book is based on fifteen years' experience in teaching statistical procedures to students of education and psychology. It may be fairly said that it is probably the best and the most up to date of the many introductory works on the subject that have so far appeared on either side of the Atlantic. As Dr. McNemar has anticipated, the influence of his own teacher, Professor T. L. Kelley, is frequently discernible, and, as a result, that of Professor Karl Pearson, under whom Professor Kelley once studied. This perhaps makes its chapters specially suitable for psychological students in this country, since so many books and papers published by British psychologists have been, and still are, profoundly indebted to Karl Pearson's contributions. At the same time, the newer methods introduced by later writers, particularly by Dr. McNemar's other teacher, Professor Hotelling, are almost equally in evidence.

The manual is designed as a one-year course for senior and graduate students. Except for two occasions on which the calculus is used, only an elementary knowledge of mathematics is presupposed. The purpose is not so much to provide the student with a knowledge of techniques adequate for original research, but rather to enable him to understand the place and use of statistical concepts and procedures in current psychology.

Three chapters are devoted to the analysis of variance and covariance. Here the presentation seems decidedly superior to that adopted in most other elementary text-books ; but its place in relation to the design of experiments might perhaps have been more clearly indicated. Further, no indication is given of the peculiar limitations of psychological data which so often prevent techniques like chi-square or analysis of variance from being taken over as they stand, and applied to problems of psychological research. Throughout the volume considerable space is given to demonstrative treatments of those topics in which the non-mathematical thinking is almost as difficult as the mathematical, for example, problems of sampling, and the notions of degrees of freedom and of interaction.

Nevertheless, the pure statistician, who has always been somewhat critical of the kind of statistical procedures favoured by psychologists, will probably have no difficulty in detecting here and there occasional inadequacies and even a little confusion in the discussion of certain logical problems. The usual distinction between large and small sample methods, which most statisticians would now consider somewhat arbitrary, is preserved. Possibly because they seem simple to the beginner, the methods of analysis appropriate to large samples are described first ; but it is not made clear that these are really limiting cases of the methods appropriate to small samples, which are statistically more fundamental. Partly as a result of this approach, the questions of sampling, estimation, and the testing of statistical hypotheses, seem at times to get confused. Thus, in the introduction, the need for methods of statistical inference is said to arise " because of certain inadequacies of research data " ; and the effect is to suggest that the major problem is one of simple estimation. The account of ' confidence ' in connexion with statistical inference is restricted almost entirely to the estimation of the population mean ; and there is no discussion of the more general problems of statistical inference as they relate to experimentation. The null hypothesis is introduced merely as " a better rule-of-thumb procedure " ; and it is elsewhere stated, perhaps a little casually, that " chi-squared can likewise be used to test the null hypothesis." The discussion of experimental design makes only an incidental appearance.

As might be expected in a book intended for psychologists, considerable attention is devoted to multiple correlation and partial regression. Doolittle's method, by no means the easiest procedure for the beginner, is recommended for solving the requisite equations when four or more variables are involved. The treatment of rank-methods, which owe their introduction into practical statistical work so largely to psychologists, is brief and somewhat out of date : its limitations are stressed, but its advantages are unmentioned, and no hint is given of the recent work of Kendall on the subject. Since discriminant functions are closely related to methods of regression, and have begun to play an increasing part in psychological research, it is perhaps a little surprising that they too are wholly omitted. Factor analysis, we are told, has been deliberately excluded, presumably because even an introductory account was considered too difficult for the beginner. Thus the special statistical procedures which have been developed for the psychologist seem to be ignored ;

and the psychological reader feels at times that the book is not so much an introductory review of *psychological* statistics as a means of introducing the student of psychology to techniques which have for the most part been developed in non-psychological fields.

The book makes a determined attempt to formulate the precise assumptions underlying the use of the various procedures as they are described, and to specify the conditions in which each method may be validly applied. This is well carried out in the discussion of correlation ; but (in the reviewer's opinion) it is less successful in the explanations of the chi-square tests. The account given of the procedure to be followed in testing the difference between variances of independent random samples is not as complete or as exact as it should be : since tables of the variance ratio are, as a rule, appropriate only for the one-tail test required in analysis of variance, it is stated that for the present case the tabled probabilities should be doubled ; but the author omits to point out that this procedure is only justifiable when the two samples are of equal size, nor indeed does he sufficiently bring out the importance of considering whether the assumption of homogeneity of variance is justified in particular cases. These, however, are blemishes which could perhaps be rectified in a later edition.

A. S.

Experimental Design in Psychological Research. By Allen L. Edwards. New York : Rinehart & Co. Inc. 1950. Pp. xvi + 446. $5.00.

Any text-book seeking to expound statistical procedures for psychological students has nowadays to keep a double aim in view. First, it should describe the more important techniques which are accepted by professional statisticians and which appear appropriate to the types of problem common in psychological research, even though such techniques have but rarely been adopted by psychologists. But secondly it must also explain those statistical methods which have in fact been in frequent use among psychologists themselves, and are therefore likely to be encountered in the books and periodicals consulted by the psychological student. If such methods have not yet been recognized by the professional statistician, or if they are now giving way to more rigorous procedures, these are not reasons for ignoring them altogether, but rather for explaining their imperfections.

This double task did not confront the earlier writers of such text-books, such as Thorndike, Kelley, or William Brown. The mathematical methods developed by Fechner and by the Galton-Pearson school were elaborated largely with psychological problems in mind. But most of the newer techniques in statistics have arisen out of work in very different fields. As mathematical devices they may be admirable in themselves ; but too often they depend, explicitly or implicitly, theoretically or practically, on assumptions which may be valid enough in the field for which they were originally intended, but are not always appropriate to the data with which the psychologist has to deal.

The point is so frequently ignored that an illustration is perhaps desirable. In a recent study of tests applied to normal and neurotic groups (the latter reclassified into various subgroups) the investigator explains that he has applied the same kind of analysis as is to be found in Snedecor's chapter on the analysis of variance. He forgets that the distinction between the normal and the neurotic is by no means so clear-cut as between ' King ' and ' Standard ' varieties of apples, and that, had the measurements furnished by his psychological tests been repeated, the reliability would have been far lower than the reliability of the figures for the weight or the quantity of fruit. The critical psychologist is therefore tempted to ask whether the differences in the data may not invalidate the methods that have thus been borrowed, and whether they do not render these more refined working methods unnecessarily laborious. How far, for example, do the classifications and the experimental procedures of the psychologist yield anything that is precisely comparable to the ' blocks,' ' plots,' and ' treatments ' which we encounter so frequently in *Statistical Methods for Research Workers ?* I do not suggest that I myself share this view. But it is, I fancy, doubts like these that have made non-statistical psychologists in this country so sceptical of the statistical approach.

Unfortunately, admirable as they are, very few of the current text-books on statistical psychology really face these difficulties. In his latest volume, Professor Edwards has fulfilled the first of the two aims I have mentioned in a way which is beyond criticism. No doubt it is the more important of the two. But the second he seems almost wholly to overlook.

As the title suggests, his general approach follows that outlined by Fisher in *The Design of Experiments*. Hence a large part of the book deals with the analysis of variance and

covariance and with the planning of factorial designs of varying degrees of complexity. A glance at the author index shows how deeply he is indebted to the work of British statisticians, like M. S. Bartlett, M. G. Kendall, F. Yates, Udny Yule, Karl Pearson, and Egon Pearson, as well as Fisher himself. And the number of actual researches on which he is able to draw in his search for concrete examples and illustrations shows how rapidly the work of such writers has begun to influence research-procedures in America. Hardly any, however, are taken from psychological experiments in this country. This is not because the work of British psychologists has been ignored, but because, until quite recently, it has so rarely adopted these newer techniques. Yet, if one turns back to some similar volume issued thirty years ago, one would find that nearly half the references relate to statistical work published by psychologists in this country. Today how many, among our senior colleagues in British psychological departments, could claim even a nodding acquaintance with " the newer developments in statistical analysis " due to British statisticians, and how many arrange for them to be taught to their research students ?

Nevertheless, I cannot help feeling that the attractive title chosen by Professor Edwards is a little too specific. Except for the fact that the examples are derived almost wholly from psychology, the title might almost as well have been phrased ' Experimental Design in *Statistical* Research.' Comparatively few of the methods described are characteristic of the typical or commoner kinds of experiment carried out in the psychological laboratory. The special statistical methods developed and used so persistently by psychologists themselves—for example, the constant process, the ranking method, the so-called discriminant methods, and factor analysis in its various branches—are not mentioned at all. Formerly statisticians were apt to retort, with Professor Filon, that " such devices are so amateurish that they should be discouraged rather than encouraged." Yet Kendall has shown that ranking procedures can be established on a sound basis ; Cramér's text-book includes references, not only to Pearson's method of principal components, but also to factor analysis by name ; while Fisher's own books include discussions of discriminant functions and the treatment of the sigmoid response curve by the probit method (i.e., what psychologists know as the constant process).

I think, too, it is a little unfortunate that the treatment might at first seem to imply that, if a psychological research is to be well planned, it must of necessity be statistical. The general principles of scientific methodology and of inductive argument are the same for all kinds of research ; and hence, except where a separate treatment is unavoidable on practical grounds, it is in my opinion a mistake, at any rate in text-books for the student of psychology, to divorce the designing of statistical investigations from the designing of scientific enquiries generally, whether statistical, or purely experimental, or merely observational.

However, it is scarcely fair to criticize an author for not discussing topics which he did not intend to take up. Professor Edwards may well reply that, at any rate for American students, these other matters are already sufficiently covered, and that his special purpose was to fill a gap that still needed filling. In his preface he explains that his primary object was " to present to the student in psychology, education, sociology, and the behavioural sciences some of the newer developments in statistical analysis, particularly with respect to small sample theory as they relate to the problems of research and experimentation in these fields." If this is taken to be his purpose, then it will be agreed that it has been excellently carried out. For students in this country who have already assimilated the older and more elementary techniques, perhaps by the aid of Professor Edwards' previous volume, there could be no better introduction to the " newer techniques." *C. B.*

The Backward Child. By Cyril Burt. London : University of London Press. Third Edition, 1950. Pp. xx + 704. 25s.

Sir Cyril Burt's book is concerned primarily with the causes and the treatment of backwardness among school children ; but it provides at the same time an admirable introduction to the study of individual differences among children of school age. Its chapters present in systematic form the results of a long series of investigations carried out by himself and by teachers working under his direction while he was Psychologist to the London County Council. To meet post-war changes, and to incorporate further knowledge accumulated during the last fifteen years, the whole volume has now been carefully revised, and the statistical sections considerably expanded.

In his earlier work on *The Young Delinquent* two new methodological features were introduced. The first consisted in an attempt to combine the case-study method with the statistical : the intention here was to use the quantitative approach to check the qualitative, and the qualitative to interpret the quantitative. The second consisted in the decision to carry out a parallel study on a control group, side by side with the study of the crucial group. The same two principles have been adopted in the investigation of educational backwardness.

In both cases the scheme underlying the examination of each child follows the same comprehensive lines, and is itself based on the results of prior statistical studies carried out by factorial methods. This is the portion of the work that has most frequently excited criticism, particularly from those who favour a different view of ' the factors of the mind.' Burt has always held that the apparent occurrence of both general and special disabilities among school children requires a fairly complex conception of mental processes—one which will include group factors as well as general factors. Not unnaturally those who set great store by " economy in psychological assumptions " were tempted to reply that one type of factor should surely be sufficient. To teachers outside training colleges and to practical psychologists concerned with the actual classification of subnormal children any such simplified explanations seemed scarcely workable. There was, however, an urgent need for a set of crucial enquiries, carried out by someone equally familiar with theoretical and with practical needs, along lines which would decide between the chief alternatives.

In this country, at the time when the volume was first published, Spearman's two-factor theory was still dominant. With a few dubious exceptions, so it was maintained, a single general ability would account for all individual differences ; and the teacher's notion of special disabilities was dismissed as an out-of-date relic of the old faculty theory, which both general and statistical psychology had at length discarded. On the other hand, in America Thurstone and his followers had introduced the opposite picture : the hypothesis of a general factor was rejected, and the existence of group factors alone admitted. A somewhat similar standpoint had previously been suggested by Thomson.

The appointment of an official psychologist working in the London schools enabled researches to be planned and carried out on a scale that would have been quite impossible for the isolated investigator. The wide and varied data thus brought together in Burt's book appears to provide a conclusive answer to the issues raised : both the statistical analyses and the examination of numerous case-histories render the double hypothesis indispensable. We seem then bound to assume first an innate general ability, resembling what is commonly called intelligence, and secondly a number of innate abilities and disabilities in particular cognitive processes, such as memory, verbal capacity, and the like. The fact that, despite attacks from two opposite fronts, this view has now won increasing acceptance points plainly to the great methodological value of joining statistical with non-statistical modes of approach in all researches of this kind. Certainly, if one may judge from its fruitfulness in the present volume, Burt's hierarchical scheme of factors, including as it does both cognitive and orectic tendencies, with the broader or more general factors ramifying into narrower, lends itself admirably to the elucidation of the problems of individual psychology, at least so far as they arise in work among school children.

In discussing the place of quantitative methods in investigations of this kind, the author insists that statistical psychology is not to be treated as a distinct or self-contained branch : the use of quantitative procedures is " merely an extension or refinement of the logical techniques which are essential to all scientific research and argument." In the main, therefore, mathematical technicalities have been kept in the background ; and in the text the critical ratios, tetrachoric correlations, and factor-saturations are merely cited incidentally in the course of the general argument to confirm or disprove the various alternative hypotheses.

During recent years, however, an increasing number of educationists, schoolmasters, and psychologists at child guidance centres have been undertaking statistical investigations on educational backwardness and kindred problems. Accordingly, in the present edition, the original appendices have been enlarged to supply a somewhat fuller account of the more useful computing devices and formulæ, and detailed references to recent publications have been added. But what should appeal to the statistical investigator most of all is the long and suggestive list of problems still awaiting systematic research. *P. B.*

INDEX TO VOLUME III

INDEX OF AUTHORS

Authors	*Titles*	*Page*
Bartlett, M. S.	Tests of Significance in Factor Analysis	77–85
Burt, Cyril	Group Factor Analysis	40–75
Burt, Cyril	The Influence of Differential Weighting	105–125
Burt, Cyril	A Reply to Sir Godfrey Thomson's Note	127–128
Burt, Cyril	The Factorial Analysis of Qualitative Data	166–185
Henrysson, S.	The Significance of Factor Loadings	159–165
Lawley, D. N.	A Method of Standardizing Group-Tests	86–89
Lawley, D. N.	Factor Analysis by Maximum Likelihood : A Correction	76
Lubin, A.	Linear and Non-Linear Discriminating Functions	90–104
Lubin, A.	A Note on Sheldon's Table of Correlations between Temperamental Traits	186–189
McLeish, J.	The Validation of Seashore's Measures of Musical Talent by Factorial Methods	129–140
Raath, M. J., and Reyburn, H. A.	Primary Factors of Personality	150–158
Reid, D. D.	Night Visual Capacity and its Relation to Survival in Operational Flying	141–149
Reyburn, H. A., and Raath, M. J.	Primary Factors of Personality	150–158
Richardson, L. F.	A Method for Computing Principal Axes	16–20
Sen, Amya	A Statistical Study of the Rorschach Test	21–39
Thomson, Godfrey	Note on Sir Cyril Burt's Paper on Differential Weighting	126
Vernon, P. E.	An Application of Factorial Analysis to the Study of Test Items	1–15

INDEX OF SUBJECTS

Titles	*Authors*	*Page*
Differential Weighting, The Influence of	Burt, Cyril	105–125
Differential Weighting, Note on Sir Cyril Burt's Paper on	Thomson, Godfrey	126
Differential Weighting, A Reply to Sir Godfrey Thomson's Note	Burt, Cyril	127–128
Discriminating Functions, Linear and Non-Linear	Lubin, A.	90–104
Factor Analysis by Maximum Likelihood : A Correction	Lawley, D. N.	76
Factor Analysis, Tests of Significance in	Bartlett, M. S.	77–85
Factor Loadings, The Significance of	Henrysson, S.	159–165
Group Factor Analysis	Burt, Cyril	40–75
Group Tests, A Method of Standardizing	Lawley, D. N.	86–89
Night Visual Capacity and its Relation to Survival in Operational Flying	Reid, D. D.	141–149
Personality, Primary Factors of	Reyburn, H. A., and Raath, M. J.	150–158
Principal Axes, A Method for Computing	Richardson, L. F.	16–20
Qualitative Data, The Factorial Analysis of	Burt, Cyril	166–185
Rorschach Test, A Statistical Study of the	Sen, Amya	21–39
Seashore's Measures of Musical Talent, The Validation of, by Factorial Methods	McLeish, J.	129–140
Sheldon's Table of Correlations between Temperamental Traits, A Note on	Lubin, A.	186–189
Test Items, An Application of Factorial Analysis to the Study of	Vernon, P. E.	1–15

BOOKS REVIEWED

Author	*Title*	*Page*
Burt, C.	The Backward Child	194–195
Edwards, Allen L.	Experimental Design in Psychological Research	193–194
McNemar, Q.	Psychological Statistics	192–193
Spearman, C., and Jones, Ll. Wynn	Human Ability	190–191

Vol. III Part I March 1950

THE BRITISH JOURNAL OF PSYCHOLOGY

Statistical Section

EDITED BY

CYRIL BURT AND GODFREY THOMSON

WITH THE ASSISTANCE OF

A. C. AITKEN
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
B. BABINGTON SMITH
W. STEPHENSON
P. E. VERNON

Managing Editor: H. G. MAULE

UNIVERSITY OF
DO

UNIVERSITY OF LONDON PRESS LTD.

WARWICK SQUARE, LONDON, E.C.4.

Price Fifteen Shillings

CONTENTS

		Page
VERNON, P. E.	An Application of Factorial Analysis to the Study of Test Items	1
RICHARDSON, L. F.	A Method for Computing Principal Axes . . .	16
SEN, AMYA	A Statistical Study of the Rorschach Test . . .	21
BURT, CYRIL	Group Factor Analysis	40
LAWLEY, D. N.	Factor Analysis by Maximum Likelihood : A Correction	76

The British Journal of Psychology (*Statistical Section*) is issued by the British Psychological Society. Three numbers will be published annually.

The subscription price is 30*s.* a volume of about 200 pages. Subscriptions should be sent to any Bookseller, or to THE MANAGER, University of London Press, Ltd., Warwick Square, London, E.C.4.

Papers for publication should be sent either to Professor Sir CYRIL BURT, Psychological Department, University College, London, W.C.1, or to Professor Sir GODFREY THOMSON, Moray House, Edinburgh, 8. Contributors will receive 20 copies of their papers free : extra copies may be had at cost price, if the order is given when the proofs are returned.

Authors are asked to send their papers in a form which is ready for submission to the printer : that is to say, the arrangement of the manuscript (including a brief list of contents at the head and a summary of results and conclusions at the end) should follow the general arrangement observed in the publications of the British Psychological Society.

Books for review should be sent to one of the Editors ; advertisements for insertion to the Manager, University of London Press, Ltd.

The British Psychological Society also issues quarterly *The British Journal of Psychology*, *General Section*, and *The British Journal of Medical Psychology*. Subscriptions, 40*s.* per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society also publishes, jointly with the Training College Association, *The British Journal of Educational Psychology*. Subscriptions, £1 per annum, for this Journal should be sent to Methuen & Co., Ltd., 36 Essex Street, London, W.C.2.

Papers for publication in the *General Section* should be sent to Professor D. W. HARDING, the Psychological Laboratory, Bedford College, London, N.W.1 ; those for publication in *The British Journal of Medical Psychology* should be sent to Dr. J. D. SUTHERLAND, 2 Beaumont Street, London, W.1 ; those for publication in *The British Journal of Educational Psychology* to Emeritus Professor C. W. VALENTINE, The White House, Wythall, Nr. Birmingham.

1956 1